PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12Q 1/68, C12P 19/34, C12N 15/00

A1

(11) International Publication Number: WO 94/29486

(43) International Publication Date: 22 December 1994 (22.12.94)

(21) International Application Number: PCT/US94/06810

(22) International Filing Date: 15 June 1994 (15.06.94)

(30) Priority Data:
08/078,471 15 June 1993 (15.06.93) US
08/117,952 7 September 1993 (07.09.93) US

(71) Applicant: THE SALK INSTITUTE FOR BIOLOGICAL STUDIES [US/US]; 10010 North Torrey Pines Road, La Jolla, CA 90037 (US).

(72) Inventors: EVANS, Glen, A.; 1652 Via Cancion, San Marcos, CA 92069 (US). SMITH, Michael, W.; 4755 Diane Avenue, San Diego, CA 92117-2905 (US).

(74) Agent: REITER, Stephen, E.; Pretty, Schroeder, Brueggemann & Clark, Suite 2000, 444 South Flower Street, Los Angeles, CA 90071 (US).

(81) Designated States: CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: METHOD FOR GENERATION OF SEQUENCE SAMPLED MAPS OF COMPLEX GENOMES

(57) Abstract

The present invention relates to a rapid and powerful sequence "sequence sampled mapping" method for sequencing complex genomes. The invention method is applicable to genomic DNA, preferably mammalian chromosomes, and in a preferred embodiment, employs a "bottom-up" mapping strategy, which allows for the simultaneous analysis of multiple cosmid clones for the detection of overlaps. The sequence sample mapping method is useful first, for the completion of high density sequence-based maps, and ultimately, for the complete sequencing of genomic DNA directly from cosmid clones.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

METHOD FOR GENERATION OF SEQUENCE SAMPLED MAPS OF COMPLEX GENOMES

The present invention relates to recombinant DNA technology. More particularly, the invention concerns a process for rapidly generating a physical sequence map of large complex genomes, including human chromosomes. The sequence mapping process ("sequence sampled mapping") depends on the use of cosmid vectors containing endogenous bacteriophage promoters to allow for the sequencing of end-specific nucleotides of each member of a contiguous library of cosmid clones.

Background of the Invention

The complete analysis of large complex genomes, such as genomes of higher eukaryotes, including human, requires the extensive isolation, purification and analysis of large fragments of DNA by cloning, generally in *E. coli*. In the past, the lambda bacteriophage cloning system has been used most frequently to generate genomic libraries. The lambda bacteriophage vectors usually accommodate inserts up to about 20 kb. Presently the primary system used to clone and manipulate large DNA fragments is that of cosmid vectors. Cosmid vectors allow the packaging of DNA fragments of up to about 45 kb in plasmids containing bacteriophage cos sites for *in vitro* packaging.

The analysis of complex genomes involves the application of both "top-down" and "bottom-up" mapping strategies. The "top-down" strategy depends on the separation on pulsed field gels of large DNA fragments generated using rare restriction endonucleases for physical linkage of DNA markers and the construction of long-range maps [Schwartz et al., Cell 37:67 (1984); Southern et al., Nucleic Acids Res. 15:5925 (1987); Burke et al., Science 236:806 (1987)]. The "bottom-up"

strategy depends on identifying overlapping sequences in a large number of randomly selected bacteriophage or cosmid clones by unique restriction enzyme "fingerprinting" and their assembly into overlapping sets of clones. "Top down" mapping is inherently more rapid and less labor intensive, but does not generate sets of DNA clones for further structural or biological analysis. "Bottom-up" mapping generates the required sets of overlapping clones but application of current strategies and pattern matching algorithms to mammalian genomes will require the analysis of thousands to tens of thousands of individual clones for the generation of complete maps.

Clone-based physical maps have been extremely useful as the framework for many types of structural and biological studies and have been constructed for several model organisms including *E. coli*, *C. elegans*, *D. melanogaster* and *S. cerevesiae* (Kohara et al., 1989, Cell, 50:495-508; Oliver et al., 1992, Nature, 357:38-46; Sulston et al., 1992, Nature, 356:37-41; Merriam et al., 1991, Science, 254:221-225). In the past few years, a variety of techniques have been utilized for the construction of ordered clone maps including cosmid and phage contig-building (Olson et al., 1986, Proc. Natl. Acad. Sci., USA, 83:7826-7830; Coulson et al., 1986, Proc. Natl. Acad. Sci., USA, 83:7821-7825), analysis of arrayed libraries (Evans and Lewis, 1989, Proc. Natl. Acad. Sci., USA, 86:5030-5034), linking libraries and pulsed-field gel analysis (Poustka and Lehrach, 1986, Trends Genetics, 2:174-179; Hermanson et al., 1992, Genomics, 13:134-143) and the assembly of YAC clone contigs (Bellanne-Chantelot et al., 1992, Cell, 70:1059-1068; Foote et al., 1992, Science 258:60-66).

Another approach for the assembly of clone maps is the use of sequence tagged sites (STSs; Olson et al, 1989, Science, 245:1434-1435): mapped DNA sequence

fragments that can be detected by amplification of specific products using the polymerase chain reaction (PCR; Saiki et al., 1988, Science, 239:487-491). STS content mapping, the analysis of STS markers present in large contiguous inserts in yeast artificial chromosomes, has proven to be an efficient method for assembling clone maps of 100 to 300 kb average resolution and has been successful for the assembly of low resolution maps of human chromosomes Y and 21 (Foote et al, 1992, supra; Chumakov et al., 1992, Nature, 359:380-387). The analysis of STS content in large DNA fragments carried in somatic cell (Delattre et al., 1991, Genomics, 9:721-727) or radiation-reduced cell hybrid lines (Cox et al., 1990, Science, 250, 245-250) also provides a powerful mapping technique.

Large numbers of mapped STS markers have been isolated for several human chromosomes (Tanigami et al., 1992, Am J Hum Genet, 50:56-64; Hori et al., 1992, Genomics, 13:129-133; Heding et al., 1992, Genomics, 13:89-94) but do not necessarily provide the needed resources for large scale chromosome mapping. In many cases, the value of these reagents is limited because the probes are poorly characterized, not generally available to the scientific community or can not be used for routine screening under a set of standardized conditions. Thus, methods of producing DNA markers suitable as reagents for large scale chromosome mapping are desired.

In addition, a major challenge of the human genome project is development of new approaches for physical analysis and sequence determination. Major progress has been made with the sequencing of large regions of DNA for significant portions of the *E. coli* genome, chromosome III of *S. cerevisiae*, several cosmid sized pieces of DNA from C. *elegans*, human, and a 100 kb T cell receptor region from mouse. See, e.g., Daniels et

al., Science, 257:771-778 (1992); Martin-Gallardo et al., Nat. Genet. 1:34-39 (1992); Oliver et al., Nature, 357:38-46 (1992); Sulston et al., Nature, 356:37-41 (1992); and Wilson et al., Genomics, 13:1198-1208 (1992). The precise determination of each base for these sequences, however, has been a labor intensive and costly undertaking.

In addition, the construction of high resolution physical maps and the acquisition of sequence have previously been considered separate efforts. Thus, new methods are desired, such as combining the steps of physical mapping and sequencing, to sequence 25%-100% of a particular portion of genomic DNA from megabase sized regions to whole genomes or chromosomes more economically than previous efforts and with reasonable accuracy.

Summary of the Invention

The present invention relates to a rapid and powerful sequence mapping method, called "sequence sampled mapping", for sequencing complex genome, said method comprising sequencing the end-specific nucleotides of each member of a library of cosmid clones, and assembling a sequence sampled map by correlating the end-specific sequence information with the relative spatial relationship between the cosmids. The invention method is applicable to genomic DNA, preferably mammalian chromosomes, and in a preferred embodiment, employs a "bottom-up" mapping strategy, which allows for the simultaneous analysis of multiple cosmid clones for the detection of overlaps. The sequence sampled mapping method permits sequence overlaps to be determined by map positions, reducing the reliance on determining regions of unique shared sequence.

In a particular embodiment of the invention, the method comprises, in any order, grouping of cosmid clones by chromosome or YAC hybridization, construction of high density cosmid contigs by restriction based fingerprinting, and direct and automated DNA sequencing from cosmid clones.

The sequence sampled mapping method is useful for the completion of high density sequence-based maps, and ultimately, for the complete sequencing of genomic DNA directly from cosmid clones. In addition, the resulting sequence information allows the detection of many genes by sequence analysis with computer programs such as FASTA, BLAST, GRAIL and others under development; allows the development of sequence tag sites (STSs) and polymorphic repeats at an actual physical spacing of a few kilobases (see, e.g., Olson et al., 1989, Science, 245:1434-1435); and allows direct PCR amplification of any part of the genome, independent of clone libraries. The invention method is also amenable to automation using the particular characteristics of the sCOS vector and cloning system. The resulting sequence sampled map is also useful, employing on-line parallel processing microcomputers which use existing software programs that have been adapted for parallel processing, for the computer analysis of genomic DNA.

Brief Description of Drawings

Figure 1 shows the vector sCOS-1 designed for cosmid multiplex analysis. The vector contains bacteriophage T3 and T7 promoters flanking a unique *Bam*HI cloning site, *Not*I sites for expedited restriction mapping and excision of the insert DNA, duplicated cos sites for high efficiency microcloning, a dominant selection for transfection into mammalian cells, Amp and Kn resistance genes, and ColE1 origin of replication.

Figure 2 illustrates the construction of cosmid vector sCOS-1. Relevant restriction sites in the precursor molecules are shown. ClaI-SalI and ClaI-XhoI fragments were excised from pWE15 and pDVcos143 and purified on agarose gels. The indicated fragments were joined using T4 DNA ligase and co-ligation of the XhoI and SalI sites resulted in the loss of both sites in the resulting plasmids.

Figure 3 depicts the DNA sequences of the cloning site, bacteriophage promoters and flanking restriction sites in sCOS vectors. Restriction sites and T3 and T7 promoter sequences added using synthetic oligonucleotides are shown. *Sfi*I, *Not*I, *Eco*RI and *Sac*II restriction sites are indicated by thin lines. The direction of transcription using T3 or T7 polymerase is indicated by the arrows and the thick lines delineate the critical nucleotides for promoter activity. The *Bam*HI site is the cloning site into which MboI-digested genomic DNA is inserted. All linkers were inserted by "linker-tailing" into the sites formed by digestion of sCOS-1 with *Eco*RI.

Figure 4 illustrates a strategy useful for analysis of physical linkage using groups of cosmids. Figure 4A illustrates that cosmids prepared in vector sCOS-1 or one of its derivatives can be used to synthesize end-specific sequences (e.g., probes for the detection of overlaps).

Figure 4B illustrates the inoculation of cosmid clones on the surface of a nitrocellulose or nylon filter from 96-well archive plates stored at -70°C. Each clone on the "grid" is assigned a unique identifying Y and X axis coordinate. Individual clones in the collection contain the innate capacity of generating probes specific for the extreme ends of the genomic DNA insert and

detecting overlapping clones on the filter. The arrows show the locations of potential overlapping clones detected by hybridization of probes generated from the clone at coordinates Y = 2, X = 7.

Figures 4C and 4D illustrate the analysis of multiple clones simultaneously. Cosmids are pooled according to the rows and columns or the matrix, DNA prepared and a mixed RNA probe synthesized. When hybridized to the matrix filter, the probe detects a pattern of spots consisting of all of the template clones and the collection of clones overlapping with one end of each of the template clones. A similar procedure is carried out using cosmids pooled according to columns of the matrix. When the two data sets are compared, hybridizing clones identified by both of the mixed probes may be overlapping with the template clone common to both sets: that clone located at the intersection of the row and column. This procedure may then be repeated using other combinations of pooled probes and either T7 or T3 polymerase. The arrows denote the location of a clone which overlaps with the "T7 end" of the clone at coordinates Y = 2, X = 4.

Figure 5 shows predicted contigs from human chromosome 11q and restriction enzyme digestion analysis. Figure 5A presents the predicted linkage and orientation of a representative cosmid contig generated by multiplex analysis of the chromosome 11q cosmid set and data analysis using the computer program "Contig-maker". The computer output indicates the coordinates of linked clones (X,Y) and the arrows denote the orientation of the linkage.

Figure 5B presents a restriction map and the location of probes used to establish unequivocal overlap of the cosmids. A restriction map of the overlapping

clones detected in Figure 5A was determined by the analysis of partial *Eco*RI digestion products hybridized with ^{32}P-labeled T3 or T7 promoter-specific oligonucleotides. Overlapping areas not confirmed by restriction map analysis were confirmed by hybridization analysis using end-specific RNA probes generated from individual cosmid clones. Cosmid clones c14,23 and c19,27 are identical. ☐ indicates bacteriophage T3 promoter, ■ indicates bacteriophage T7 promoter.

Figure 6 presents physical and sequence maps of the (A) β-giardin and (B) a second random genomic region with arrows at the ends of cosmids corresponding to their t7 ends. Smaller hash marks represent locally regionalized, but not fully ordered restriction fragments. Regions of sequence which match the genes annotated on the figure are shown with a greyed-in box.

Figures 7A-F present alignment of various regions of sequence homology found by BLAST searches of the protein sequence databanks with cosmid end sequences.

Figure 8 presents a histogram showing the distances between ordered cosmid ends and their frequency of occurrence in the two contigs determined as described herein.

Figure 9 presents a plot of the number of sequenced cosmid ends versus the total amount of *Giardia lamblia* genome sequenced. Calculations are based on the equations of Lander and Waterman [Genomics, 2:231-239 (1988)] assuming an overlap detection of 50 bases.

Figure 10 presents regional mapping of a chromosome 11 STS using a panel of somatic cell hybrids. This analysis shows the regional mapping of an STS to bin 2, FLpter 0.05-0.24, by PCR analysis and is typical of

the bulk of STS primer results derived under these conditions. The hybrid mapping panel breakpoints are shown relative to human chromosome 11.

Figure 11 presents protein sequence alignments of putative genes detected from analysis of DNA sequences. DNA sequence determined from cosmid clones were translated into six reading frames and used to search GenPept, PIR or Swiss-Prot protein sequence databases using BLASTX. The clone name from which the sequence was derived is shown next to its translated sequence. The flanking numbers indicate the matching position in the nucleotide sequence of the cosmid or the protein sequence of the database entry. The one amino acid code translation is shown with X=any amino acid (generally caused by the inability to determine a base) and *= a stop codon. The X in cSRL-7d2 has a one in four chance of being a stop codon.

Detailed Description of the Invention

In accordance with the present invention, there is provided a method for sequencing complex genomes. The invention method comprises:

(1) sequencing the end-specific nucleotides of each member of a library of cosmid clones,

wherein said cosmid clones are prepared by inserting genomic DNA fragments into cosmid vectors,

wherein the cosmid vectors include sequences of nucleotides that flank at least one end of the inserted DNA, and that serve as transcription initiation sites for the synthesis of a nucleic acid specific to the ends of the inserted DNA, and

(2) assembling a sequence sampled map by correlating the end-specific sequence information with the relative spatial relationship between the cosmids.

In a preferred embodiment, the invention sequence sampled mapping method provides for the sequencing of the entire genome of any organism for which genomic DNA is available, preferably mammalian genomic DNA, more preferably human genomic DNA.

As used herein, the phrase "end-specific nucleotides of each member of a library of cosmid clones" refers to the nucleotide sequences at the extreme 5' and 3' ends of a given genomic DNA insert. Typically, the amount of nucleotides sequenced from each end-specific nucleotide sequence will be at least 100, preferably 250, more preferably 350, yet more preferably 550, with at least 1000 nucleotides being especially preferred. The amount of sequenced nucleotides required for the practice of the invention method varies as a function of the depth of cosmids to be sequenced.

The phrase "depth of cosmids", and grammatical variations thereof, refers to the number of overlapping cosmids that contain, in common, a specified region (i.e., 1 nucleotide) of genomic DNA to be mapped and sequenced. For example, a 20X (20-fold) depth of cosmids covering a specified region of genomic DNA refers to 20 cosmids that, on average, contain at least one nucleotide of genomic DNA in common.

The depth of cosmids is chosen so as to maximize the number of unique genomic DNA insert ends and to provide a desired average spacing between the respective 5' or 3' ends of two consecutive contiguous cosmid clones. For example, if on average, each cosmid contains approximately 40 kb of genomic DNA insert, then

a cosmid depth of 20X would produce, on average, a spacing of 1000 nucleotide base pairs between the respective 5' or 3' ends of each consecutive genomic DNA insert. Thus, sequencing approximately 500 base pairs of the 5' and 3' ends of all given cosmids will provide genomic DNA sequence data for approximately 50% of the respective genomic DNA sample.

The depth of cosmids can be varied by methods well-known in the art. One method is to select any one or combination of restriction enzymes that recognize a specific genomic DNA sequence, preferably a 4-bp sequence ("4-bp-recognizing"). Next the restriction enzyme(s) are employed to either partially or completely digest a given genomic DNA sample so as to produce genomic DNA insert fragments approximately 40-45 kb in length that are unique with respect to other genomic DNA insert fragments by as little as 100 base pairs to as great as 5 - 10 kb. Restriction enzymes that recognize 4 bp sequences suitable for use herein include, for example: Sau3A, AccII, AluI, BSP50, FnuDII, HaeIII, HhaI, the isoschizomers thereof, and the like.

A twenty-fold library generated by partial digestions of genomic DNA with several different restrictions enzymes would greatly increase the number of potential cloning sites and reduce the number of cloned ends which are exactly the same. For example, multiple 5-10 fold deep libraries, from the same genomic DNA source, can each be generated with a unique four-bp-recognizing restriction enzyme. This would require straightforward modifications to the COS vectors described hereinater, e.g., adding appropriate polylinker sites. The uniquely restricted libraries can then be combined to arrive at libraries with a substantially high level of cosmid depth (e.g., at least about 20-fold deep, preferably at least about 40-fold deep, and more

preferably at least about 50-fold deep). Thus, a 21-fold library may be constructed from three sub-libraries of 7-fold cosmid depth, whereby each sub-library is produced with a different four-bp-recognizing restriction enzyme.

One of skill in the art will recognize that by varying any one or both of the above-described parameters, sequencing of varying percentages of a given genomic sample becomes feasible, such as at least 25%, preferably 50%, more preferably 75%, with 100% (i.e., the entire genome) being especially preferred (see, e.g., Figure 9). For example, in the above-example, increasing the depth of cosmids from 20X to 40X and sequencing 500 bp of each end of genomic DNA inserts provides the sequence for 100% of the 40 kb genomic DNA sample. Alternatively, in the above-example, increasing the amount of nucleotides sequenced from 500 to 1000 bp also provides the sequence for 100% of the 40 kb genomic DNA sample. Stated another way, increasing the average lengths of sequences determined, to about one kilobase, would result in nearly complete one-pass sequencing of a genome or chromosome at a fraction of the cost.

Either prior to, concurrently, or following the construction of contigs or determining the relative spatial relationship between the cosmid clones described above, the sequencing of end-specific nucleotides step of the present invention, is conducted to provide the sequence information that will be assembled into the "sequence sampled map".

The sequencing step may be carried out either manually or using an automated DNA Sequencer employing well-known methods, such as, for example, specific or degenerate primer extension, transposon primer insertion, ordered deletion, random shot gun sequencing, sequencing by hybridization, and the like. In a preferred

embodiment, the 5' and 3' ends of each cosmid clone within a cosmid library is subjected to "one pass" (i.e., sequencing only once) automated DNA sequencing as described in Examples 3 and 4. Automated DNA sequencing devices are well-known and widely available to those of skill in the art, such as, for example, the sequencing devices available from Applied Biosystems (e.g., ABI 373A Sequencer combined with a Catalyst 800 robot, Foster City, CA), Pharmacia (Piscataway, NJ), Millipore (Milford, MA), and the like.

It is recognized that automated DNA sequencing technology is currently progressing extremely rapidly. Thus, automated sequencing methods and devices that will allow sequencing of DNA fragments greater than 500 nucleotides (i.e., 1 to 5 kb) are also contemplated in the methods described herein (see, e.g., Ansorge et al., 1992, Electrophoresis, 13:616-619). Subsequently, when correction and verification of sequence information is desired for a particular region, an independent sequencing methodology may be employed, e.g., sequencing by hybridization, and the like.

Raw sequence information obtained from automated sequencing, or sequence sampled mapped sequence, can be analyzed immediately using on-line parallel processing microcomputers that employ existing software programs adapted for parallel processing. Sequence analysis software programs contemplated for use herein include, for example: GRAIL, which locates protein-coding regions in genomic DNA sequences (see, e.g., Uberbacher et al., PNAS, USA, 88:11261-11265, 1991); BLAST-n and BLAST-x, which compares sequence similarity between nucleotides and amino acid sequences, respectively (see, e.g., Altschul et al., J. Mol. Biol., 215:403-410, 1990); FASTA, which identifies sequence repeats (see, e.g., Pearson et al., PNAS, USA, 85:2444-

2448, 1988). Raw sequence information may also be used advantagously to generate PCR primers useful in PCR assays for polymorphic repeats around specific sequences of interest, e.g., around "CA" nucleotide runs, other simple sequences, and the like.

In another aspect of the invention, prior to the completion of a complete physical map, raw sequence information can be used to generate "sequence-tagged sites" (STSs) as described in Example 4. The STSs can be used, e.g., for producing an ordered set of YACs, for the analysis of sites of chromosomal pathology (e.g., translocations, polymorphic repeats, and inversions), and the like. The production of STSs allows access to mapping markers based upon PCR amplification of known genomic sequence.

Briefly, as described above, DNA sequences are determined by sequencing directly from cosmid templates using primers complementary to the promoters (e.g., T3 and T7) present in the cloning vector. Oligonucleotide PCR primers are predicted by computer from a suitable amount of randomly selected cosmid-end-derived sequences, and are tested using a battery of genomic DNA templates, preferably corresponding to a specific chromosome. Cosmids are then regionally localized to the respective chromosome using fluorescence *in situ* hybridization and/or by the analysis of a somatic cell hybrid panel. Additional STSs corresponding to known genes and genetic markers on the respective chromosome may also be produced under the same series of standardized conditions.

As used herein, a "suitable amount" of STSs produced can be varied by one of skill in the art to provide a desired coverage (preferably uniform) of a respective chromosome. For example, it is well within the skill in the art to select an amount of STSs that

provides an average spacing between conscecutive STSs within the range of about 1 kb up to about 500 kb, and also provides sufficient density for STS content mapping using YAC clones or contigs (see, e.g., Bellanne-Chaltelot et al., 1992, Cell, 70:1059-1068). Thus, e.g., assuming a chromosomal size of 126 mb (e.g., based upon chromosome 11 comprising 4.2% of the human genome), a collection of 370 STSs will have an average spacing of one STS per 340 kb and would provide sufficient density for STS content mapping using YAC clones. In a preferred embodiment, the quantity of STSs produced by the methods described herein provides an average spacing between conscecutive STSs within the range of about 1 kb up to about 5 kb.

As used herein, the phrase "assembling a sequence sampled map" refers to the step of ordering the nucleotide sequences obtained from the sequencing step into the order in which they naturally occur in the source genome. This is accomplished by correlating the end-specific sequence information obtained in step (1) above with the relative spatial relationship between the cosmids.

Once the sequence sampled map has been determined, a minimum tiling path of cosmids (just enough to cover the region once) can be used as sequencing templates. Each sampled sequence can then be extended in both directions to triple the effective sequence reads. The availability of inexpensive oligonucleotide primers will make this sequence walking an attractive option for finishing the sequence. One attractive approach which may make primer walking affordable is the use of contiguous hexamer oligonucleotides to specifically prime sequencing reactions (see, e.g., Kieleczawa et. al., 1992, Science, 258:1787-91).

The term "relative spatial relationship between the cosmids" refers to the physical mapping of the genomic DNA inserts into the contiguous order and/or the respective orientations (e.g., 5'-3' or 3'-5') in which they naturally occur in the source genome. The relative spatial relationship may be determined either prior to, concurrently, or after the step of "sequencing the end-specific sequences of each member of a library of cosmid clones." In a preferred embodiment, the relative spatial relationship between the cosmids is determined concurrent (i.e., in parallel) with the sequencing of step (1) above.

The relative spatial relationship between the cosmids (i.e., physical map) can be determined by methods well-known in the art, such as fingerprinting by restriction enzyme mapping ("restriction-fragment-length mapping") employing several different restriction enzymes [see, e.g., Olson et al., Proc. Natl. Acad. Sci. USA 83:7826 (1986); Coulson et al., Proc. Natl. Acad. Sci. USA 83:7821 (1986); Kohara et al., Cell 50:495 (1987); and the like]; the "cosmid multiplex analysis" method as described in U.S. Patent No. 5,219,726, incorporated herein by reference in its entirety; and the like.

Restriction fragment length mapping employs full and/or partial restriction enzyme digests. The partial digestion provides the order of restriction within a cloned genomic insert, and the full digestion provides the exact sizes of the restriction fragments. Partial digestion products are detected with non-insert probes specific to either side of the cosmid vector, which orients the sequence relative to the genomic map. The partial digestion restriction maps are reconciled with data from the full restriction digest to obtain a spatially correct physical map. The restriction enzyme maps of each cosmid are compiled into an overall map of

the genomic region of interest, which could be derived from a YAC (Yeast Artificial Chromosome), BAC (Bacterial Artificial Chromosome), PAC (P1 Artificial Chromosome), MAC (Mammalian Artificial Chromosome), a whole chromosome, or a small whole genome.

In another embodiment of the present invention, the relative spatial relationship between the cosmids is determined by the "cosmid multiplex analysis" method as described in Example 2 and U.S. Patent No. 5,219,726. The cosmid multiplex analysis method depends on the use of cosmid vectors allowing for the synthesis of corresponding RNA sequences (probes) or DNA sequences specific to the extreme ends of the DNA fragments contained within the cosmid, directly from the DNA inserts of the cosmid.

Briefly, cosmid libraries are constructed using vectors containing at least one bacteriophage promoter adjacent to the genomic DNA insert, positioned operatively for the transcription thereof. Preferably, the cosmid vectors contain two bacteriophage promoters flanking the DNA fragment ligated into the insertion site. Synthesis of an end-specific RNA probe from any clone in the collection allows the overlapping clones to be easily detected by hybridization. Because this strategy does not depend on pattern recognition for detecting overlaps, analysis may be carried out simultaneously on cosmid clones grouped together.

The "cosmid multiplex analysis" method is suitable for the unambiguous detection of overlapping regions as small as several hundred nucleotides in contiguous cosmids. Accordingly, the number of clones needed for map closure can be reduced by up to three-fold. Finally, this strategy represents essentially simultaneous cosmid "walking" and thus is

basically non-random, allowing the investigator the freedom to pause and investigate some interesting biology rather than requiring completion of the map before it becomes useful.

It has been found that significant improvements in the speed and efficiency of "bottom-up" genomic mapping can be achieved, by 1) isolating restricted regions of large mammalian genomes in a "sublibrary" preorganized on a solid matrix, 2) using hybridization of end-specific probes for detection of overlapping clones in the collection, and 3) analyzing multiple clones simultaneously for the detection of all overlaps in the collection.

In accordance with the present invention, direct sequencing can then be carried out on the fragment ends of the individual cosmids which make up the resulting contig. This will generate in the range of about 350-550 base pairs of sequence information, separated by gaps of about 1-5 kb, depending on the chosen depth of cosmids employed. Thus, for a cosmid depth of 10-20X, in the range of about 20-50% of the complete chromosome sequence can be obtained very rapidly and at relatively low cost.

Thus, an alternate embodiment of the present invention provides a method for sequencing complex genomes, said method comprising:

(1) preparing a genomic library of cosmid clones by inserting DNA fragments from said genome into cosmid vectors, wherein the cosmid vectors include sequences of nucleotides that flank at least one end of the inserted DNA, and that serve as transcription initiation sites for the synthesis of end-specific probes,

(2) arranging the cosmid clones, whereby each clone may be identified and replicas of said arrangement may be reproduced,

(3) pooling portions of said cosmid clones and synthesizing pools of mixed end-specific probes from the DNA inserts that have been prepared from said pooled clones, wherein each pool contains fewer than all of the cosmid clones in the library, but all of the cosmid clones in the library are included in at least one pool,

(4) hybridizing each pool of probes to a replica of said arranged cosmid clones and identifying the cosmid clones in each replica that hybridize to the probes, wherein said identified clones include the pooled cosmid clones and cosmid clones that contain DNA inserts that overlap with the DNA inserts in the pooled clones,

(5) identifying the cosmid clones from among those identified in step (4) that hybridize to two or more pools of probes, thereby identifying groups of cosmid clones that include overlapping DNA,

(6) assembling contigs from said groups, and

(7) sequencing the end-specific nucleotides of each overlapping member of said cosmid clones.

In a preferred embodiment, the cross-hybridizing clones are identified by pairwise comparison of data sets obtained from two groups of cosmid clones containing at least one common clone. The cosmid clones are preferably pooled according to the rows and columns of a two-dimensional matrix.

Preferably, the cosmid vectors used in the above processes comprise two oppositely oriented promoters, each of which is specific for a bacteriophage RNA polymerase, positioned on two sides of the cloning site. Most preferably, the vectors contain T3 and T7 endogenous bacteriophage promoters flanking the cloned genomic DNA. Vectors containing at least two cos sites

20

are particularly preferred, since they allow the use of DNA fragments without the need for prior size separation.

From the list of linked clones produced by this technique, contigs can be assembled either manually or through computer analysis of the data, then the fragment ends of the DNA inserts in the individual cosmids can be subjected to automated DNA sequencing. This will typically generate ordered sequence fragments about 350-550 base pairs in length, separated by gaps in the range of about 1-5 kb, depending on the chosen cosmid depth.

As used herein, the term "genomic library" refers to a mixture of clones constructed by inserting fragments of genomic DNA into a suitable vector. The term "library" implies the existence of large numbers of different recombinants out of which only a few are of immediate interest to the investigator. Genomic DNA can be the entire genome, a single chromosome, or a portion of a chromosome, such as the 300-400 kb portions of chromosomal DNA typically contained within YACs, of a given organism.

The terms "cosmid" and "cosmid vector" and grammatical variations thereof, are used interchangeably herein and refer to plasmid vectors that contain a lambda bacteriophage cos (cohesive end) site. The lambda bacteriophage packaging system selects DNA molecules of about the size of the lambda genome (37-52 kb). Accordingly, plasmid recombinant DNA having a minimum size of about 38 kb and a maximum size of about 52 kb (about 78% and about 105% of phage lambda, respectively), can be packaged *in vitro* in the lambda phage coat. In addition to the cos site(s) cosmid vectors usually contain a marker gene allowing for selection in bacteria (antibiotic resistance gene), and one or more unique

restriction sites for cloning. Plasmids with a large variety of cloning sites and prokaryotic and eukaryotic selection markers can be converted to cosmids by insertion of the lambda cos region.

The term "cosmid clone" refers to a cosmid vector that contains a genomic DNA insert. The term "plasmid" refers to circular, double-stranded DNA loops which in their vector form, are not bound to the chromosome. The term "nucleic acid" refers to a synthetic or naturally occurring DNA or RNA molecule.

As used herein, the term "a promoter specific for a bacteriophage RNA polymerase" means a wild-type or non-wild-type promoter that can be used by the bacteriophage RNA polymerase for *in vitro* transcription of a DNA fragment. When a non-wild-type promoter is used for such *in vitro* transcription of a DNA fragment, transcription will occur at a rate which is at least 10% of the rate at which transcription would have occurred if a wild-type or native promoter had been used by the bacteriophage RNA polymerase to transcribe the DNA fragment *in vitro*.

The term "cloning site" as used herein, means restriction endonuclease site on the DNA sequence of the cosmid vectors of the present invention where a DNA fragment can be inserted without deleting any of the original DNA.

Reference to a promoter positioned "operatively for transcription of a DNA fragment", as used herein, means that the promoter will be positioned in such a way that any DNA sequences between the promoter's transcriptional start site and the DNA fragment will not prevent transcription of at least a portion of the DNA fragment by the promoter. The term "at least a portion"

means that preferably at least 8 base pairs and more preferably at least about 30 bp of the DNA fragment will be transcribed.

The terms "end-specific RNA sequences", "RNA probes", and grammatical variations thereof, are used to refer to hybridization probes obtained by transcription of corresponding DNA fragments.

Clones are overlapping if they contain contiguous DNA in the same relationship as that in the genome. One method for detecting overlaps is to synthesize an RNA probe from one end of a first clone. If this probe detectably hybridizes with an end of the second clone under standard hybridization conditions, the two clones are overlapping [Wahl et al., PNAS USA 84:2160 (1987)].

The term "contig" was introduced by Rodger Staden, Nucleic Acids Res. 8:3673 (1980) in connection with DNA sequence analysis, and refers to groups of clones with contiguous nucleotide sequences.

The cosmid multiplex analysis method employs essentially the strategy illustrated in Figure 4 used for genomic mapping using cosmid vectors.

In a first step, a genomic library which represents a limited portion of a genome is constructed in a cosmid vector allowing for the synthesis of RNA probes and DNA strands for sequencing directly from insert DNA using endogenous bacteriophage promoters. A convenient and powerful way of subdividing the human genome for the preparation of libraries is through chromosome purification by flow cytometry [Gray et al., Cold Spring Harbor Symp. LI 1986 p. 141].

Cosmid vectors suitable for constructing a genomic "sub-library" include the pWE vectors described by Wahl et al., in Proc. Natl. Acad. Sci. USA 84:2160 (1987), e.g. pWE2, pWE4, pWE8 pWE10, pWE15, and pWE16, preferably pWE15 and pWE16. The construction of these vectors is described in the Materials and Methods section of the cited article and in Evans et al., Methods in Enzymology 152:604 (1987). These vectors, in addition to replication and selection functions, such as plasmid origin of replication, bacterial genes specifying antibiotic resistance, and the bacteriophage lambda cohesive termini (cos sequences), contain the transcription promoters from either bacteriophage SP6, T7 or T3 flanking a unique *Bam*HI cloning site.

In one embodiment, cosmid vectors containing a duplicated cos sequence are employed. These "sCOS" vectors have the following important characteristics: 1) the presence of two cos sites such that packaging could be carried out with high efficiency and without requiring size selection of the insert DNA; 2) the presence of T3 and T7 bacteriophage promoters for the synthesis of "walking" probes; 3) unique restriction sites for removing the insert and to aid in restriction mapping; 4) selectable genes for gene transfer in eukaryotic cells; and 5) a plasmid origin of replication giving a high yield of cosmid DNA for preparing templates.

The construction of plasmid sCOS-1 is described in the Examples, and is illustrated in Figures 1 and 2. The design of this plasmid (and derivatives thereof) allows for rapid production of RNA probes specific for both ends of the inserted DNA sequences. In addition, this design allows for the automated sequencing of the 5' and 3' ends of the genomic DNA insert.

Plasmid vector, sCOS-1, shown in Figure 1, is 6.7 kb in size and has a cloning capacity of 31 to 48 kb. As with pWE vectors, bacteriophage T3 and T7 promoters were oriented into the *Bam*HI cloning site to allow direct synthesis of end-specific RNA probes for molecular "walking" and to allow sequencing the end-specific nucleotides of a given genomic DNA insert. Previous experience with pWE cosmids suggested that *Not*I restriction sites may not be ideal for excision or mapping of inserts in some regions of the genome where *Not*I sites might be clustered. Therefore, additional cosmid vectors with other rare restriction sites have been constructed, by substituting the cloning/polymerase sites of sCOS-1 with sequences containing *Not*I and *Sac*II sites (sCOS-2) or SfiI sites (sCOS-4). The asymmetric rare sites in sCOS-2 are useful for cloning ends of large *Not*I or *Sac*II fragments for isolation of "linking" clones for long range mapping by pulsed field gel analysis [Buiting et al., Genomics 3:143 (1988)]. Also, vectors which lack *Not*I sites, such as sCOS-4, would potentially allow the selection of clones containing unique *Not*I junction fragments by hybridization with *Not*I-specific oligonucleotides [Estivill et al., Nucleic Acids Res. 15:1415 (1987)].

The double cos site sCOS vectors make feasible the preparation of representative libraries from very small amounts of purified, partially digested DNA, and are, therefore, presently preferred for carrying out the method of the present invention.

One of skill in the art can modify the sCOS vectors employed herein in a variety of ways to make them more suitable for restriction-fragment-length mapping using partial digestion (see, e.g., Kohara et al., *Cell* 50:495-508 (1987). For example, an "sCOS-derivative" vector useful in the invention methods described herein

can include, intron encoded endonuclease sites which recognize approximately 20 base pairs, and a multiple cloning site (i.e., polylinker) which will ligate with partially-digested genomic DNA from most commercially available restriction enzymes that recognize four bases and generate staggered ends.

For determining the relative spatial relationship between the cosmid clones, the individual clones of the genomic library are arranged on a nitrocellulose or nylon filter matrix and each clone is identified by unique coordinates. If the randomly chosen clones are arranged in a two-dimensional matrix, they are identified by unique X and Y coordinates. For convenience in handling, the pattern of the matrix is preferably based on the pattern and spacing of wells of a standard 96-well microtitre plate and the repetitive preparation of culture plates and hybridization filters may be carried out using equipment designed for working with this standard. Each individual cosmid clone in the collection possesses the innate mechanism of generating an RNA probe capable of detecting any overlapping or identical clones in the collection.

If an RNA probe is generated using T3 or T7 polymerase, and overlapping clones are detected by hybridization of the probe to a replica of the filter grid, using cosmid clones arranged on a 36 x 36 matrix containing 1296 clones, all of the overlaps can be detected by carrying out 1296 T3 polymerase reactions, 1296 T7 reactions and subsequent hybridization reactions.

However, as an alternative to the individual analysis of cosmid clones for the detection of overlaps, simultaneous analysis of multiple cosmid clones in groups can be conducted, as described in U.S. Patent No. 5,219,726. Accordingly, this preferred strategy allows

the analysis of a collection of cosmid clones with far less effort than "fingerprinting" each of the clones selected individually.

The linked clones detected by the above method can then be grouped into contigs, either manually or, preferably, using appropriate computer programs. To confirm the correctness of the groupings, some of the contigs can be subjected to detailed restriction enzyme analysis, and the degree of physical overlap along with a physical map can be determined. To complete a physical genomic map, the above-outlined procedure can be repeated with as many clones as necessary, and the gaps between the contigs can be filled in, e.g. by traditional chromosome walking.

The method described hereinabove represents a special case of a more general sequence mapping strategy based on clone matrices of higher order, such as, for example, greater than or equal to 3 dimensions (as described in U.S. Patent No. 5,219,726).

Table 1

Theoretical analysis of genomes of various sizes by cosmid multiplex analysis where clones are organized in matrices of various dimensions and clones analyzed using probes prepared from groups of cosmid clones.

dimension*	matrix size	probe pool size	number of clones analyzed	number of mixed probes
n	d	d^{n-1}	d^n	n x d
2	10	10	100	20
3	10	100	1000	30
2	36	36	1296	72
3	18	324	5832	54
3	36	1296	46,656	108
4	25	15,625	390,625	100
5	13	28,561	400,000	65

* Where a matrix of n dimensions containing d^n cosmids is used for the arrangement of cosmid clones, pooled probes of d^{n-1} members are prepared which allow for the analysis of d^n individual clones by multiplex analysis using n x d mixed probes. For example, using a three dimensional matrix of 10 x 10 members, pooled probes of 100 individual clones could be used to analyze 1000 individual cosmids with 30 analytical reactions.

Given that the theoretical limit of detectable signal from pooled RNA probes is not reached, cosmids representing the entire human genome with four-fold redundancy can potentially be analyzed by the "multiplex" using a 5-dimensional matrix of n=13 in only 65 hybridization reactions. The theoretical limitation of this strategy seems to be the number of individual clones which can be pooled and still give reproducible hybridization signal. Current protocols suggest that this limitation may be somewhat greater than 100 clones.

Further details of the invention are illustrated by the following, non-limiting examples.

Examples

Unless otherwise stated, the present invention was performed using standard procedures, as described, for example in Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA (1982); Sambrook et al., Molecular Cloning: A Laboratory Manual (2 ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA (1989); Davis et al., Basic Methods in Molecular Biology, Elsevier Science Publishing, Inc., New York, USA (1986); or Methods in Enzymology: Guide to Molecular Cloning Techniques Vol.152, S. L. Berger and A. R. Kimmerl Eds., Academic Press Inc., San Diego, USA (1987).

Cosmid vectors

Genomic libraries were constructed in cosmid vector sCOS-1 (illustrated in Figure 1). sCOS-1 was prepared from cosmid vectors pWE15 [see Evans et al., Methods in Enzymology 152:604 (1987)]; ATCC Accession No. 37503, American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, Maryland 20852 USA] and pDVcos134 [a gift from J. Reese, in wide circulation among scientists]. pWE15 DNA was digested with *Cla*I and *Sal*I, and the 6kb *Cla*I-*Sal*I restriction fragment, lacking the *cos* sequence was purified. Cosmid pDVcos134 was digested with *Cla*I and *Xho*I and a restriction fragment containing the duplicated *cos* region was purified on a low melting point agarose gel. The purified fragments were ligated using T4 DNA ligase and transformed into *E. coli* host strain DH5, which is a derivative of the strongly recA⁻ strain DH1 (commercially available, e.g. from Bethesda Laboratories, Gaithersburg, MD, USA). Alternatively, purified fragments can be transformed into such host cells as AG1 (Stratagene Cloning Systems, San Diego, CA), a derivative of DH5 selected for high packaging efficiency, or into HB101 (commercially available, e.g. from Bethesda Laboratories, Gaithersburg, MD, USA).

Other pWE plasmids suitable for genomic mapping according to the invention are disclosed in Evans et al., Methods in Enzymology, Supra. Cosmid vector pWE16 has been deposited with the American Type Culture Collection, and has been accorded ATCC No. 37524.

Cosmids sCOS-2 and sCOS-4 are derivatives of sCOS-1 where the cloning site has been altered to substitute other rare restriction sites for the *Not*I sites. Cosmid vector sCOS-2 was constructed by digesting sCOS-1 with *Eco*RI, and purifying the plasmid DNA away

from the *Not*I-T3 promoter-*Bam*HI-T7 promoter-*Not*I linker sequence by ethanol precipitation. A 30-nucleotide double-stranded synthetic oligomer with *Eco*RI coadhesive ends, containing *Not*I-T3 promoter-*Bam*HI-T7 promoter-*Sac*II sequences was added by linker-tailing [Lathe et al., DNA 3:173 (1984)]. sCOS-4 was constructed using a similar procedure adding a double-stranded synthetic oligonucleotide containing *Eco*RI coadhesive ends and a *Sfi*I-T3 promoter-*Bam*HI-T7 promoter-*Sfi*I sequence.

EXAMPLE 1

Construction of cosmid libraries in sCOS vectors

High molecular weight genomic DNA for cosmid cloning was prepared by proteinase k digestion and gentle phenol extraction followed by dialysis [DiLella et al., Methods in Enzymology 152:199 (1987)]. The average molecular size of the isolated DNA was determined using field inversion gel electrophoresis [Carle et al., Science 232:65 (1986)] and ranged from about 500 kb to greater than 3 mb. DNA was digested with *Mbo*I under conditions recommended by the manufacturers and the digestion terminated by phenol/chloroform extraction. Following digestion, the DNA was analyzed on field inversion gels or 0.3% agarose gels to determine the average size of the digestion products. For the construction of genomic libraries in cosmid vector sCOS-1 genomic DNA was digested to an average size of 100 - 120 kb, and dephosphorylated with calf intestinal phosphatase. The genomic DNA was not size separated before cloning.

Vector cloning arms were prepared by first digesting purified sCOS vector DNA with *Xba*I followed by dephosphorylation with calf intestinal alkaline phosphatase. The reaction was terminated by phenol/chloroform extraction and the DNA collected by

ethanol precipitation. The linearized, dephosphorylated vector DNA was then digested with *Bam*HI, extracted with phenol/chloroform and stored at a concentration of 1 mg/ml in 20 mM TRIS·HCl, pH 6, 1 mM EDTA. Ligations were performed using 1 μg of vector arms and 50 ng to 3 μg of genomic DNA. Reactions were incubated with 2 Weiss Units of T4 DNA ligase and packaged using commercial *in vitro* packaging lysates. Bacteriophage lambda packaging extracts may contain significant amounts of *Eco*K restriction activity. To avoid the possibility that mammalian sequences containing an *Eco*K site might be underrepresented in the library, genomic libraries are prepared using *in vitro* packaging extracts which lack *Eco*K restriction activity (e.g. Gigapak-Gold; Stratagene Cloning Systems, San Diego, CA).

Cosmid libraries were plated directly on LB agar (LB media containing 1.2% Bacto-agar; autoclave) containing 25 μg/ml of kanamycin sulfate and libraries screened without further amplification [Evans et al., Methods in Enzymology 152:604 (1987)]. Libraries were stored as original non-amplified plate stocks in LB media (10 g Bacto-tryptone, 5 g yeast extract, 5 g NaCl per liter of water; autoclave) with 15% glycerol at a concentration of 2.2×10^{11} bacteria/ml at -70°C. The cosmid library used in the study described in the examples consisted of 1.5×10^{7} independent clones.

Selection of Human Clones from a Somatic Cell Hybrid Genomic Library

Cosmid libraries were plated on 570 cm^2 LB agar trays at a density of 10 clones/cm^2, replica filters prepared and filters hybridized with human placenta DNA labeled with ^{32}P-dCTP to a specific activity of 10^8 cpm/μg. Under these hybridization conditions, no background hybridization was detected against cosmids

carrying mouse genomic DNA. Cosmids containing human genomic DNA inserts were picked with toothpicks, rescreened by hybridization to ^{32}P-labeled human DNA, and archived in 96-well microtitre plates containing LB media, 15% glycerol and 25 μg/ml kanamycin sulfate at -70°C. Individual clones isolated from cosmid libraries were routinely grown, replicated, and DNA prepared using standard round-bottom 96-well microtitre plates. Replica transfer of clones in 96-well microtitre plates and transfer from archived plates to screening filters was carried out using an aluminum "hedgehog" made from 3-mm diameter brass rods set in plastic block, as described by Coulson et al., Supra (p. 7822), or a laboratory robot (Beckman Biomek 1000).

Plating and Screening Libraries

For multiplex analysis, archived cosmids were inoculated on the surface of a nitrocellulose or nylon based filter in a matrix or "grid" pattern. The size and density of the "grid" was determined by the pattern of wells in a standard 96-well microtitre plate and, in the experiments described in the examples, a 36 x 36 matrix was used. Before applying bacterial culture, a matrix pattern prepared on paper was transferred directly to the filter membrane by passing the filter through a copying machine followed by autoclaving. The clones were allowed to grow on the surface of the filter at 37°C for 12 to 15 hours and bacterial DNA was fixed to the filter using a standard colony lysis procedure [Vogeli et al., Methods in Enzymology 152:407 (1987)].

RNA Probe Synthesis and Hybridization Reactions

Cosmids were transferred from archives to fresh 96-well plates containing liquid LB media with 25 μg/ml kanamycin sulfate and incubated at 37°C in a humidified

atmosphere for 6 to 10 hours. Supernatants from individual wells were pooled and DNA prepared using a previously described cosmid miniprep procedure [Evans et al., Methods in Enzymology, Supra]. Cosmids constructed with vector sCOS-1, or one of its derivatives, yield up to 2 μg of DNA from a 300 μl culture. All probe synthesis and mapping reactions were carried out with DNA prepared from minilysates. In some cases, the pooled DNA was digested with a restriction endonuclease such as *Bam*HI or *Hind*III prior to probe synthesis. RNA probes were synthesized using bacteriophage T3 or T7 polymerase (Stratagene Cloning Systems, San Diego, CA, USA). Thus, cosmid DNA was prepared and 1-2 μg of the DNA was transcribed with T7 or T3 RNA polymerase in a 20 μl reaction, as described by Melton et al. (1984) Nucleic Acids Research 12:7035-7054, using 50 μCi of [α-^{32}P] UTP and 12 μM unlabeled UTP. ^{32}P-UTP and polymerase reactions were terminated by extraction with phenol and chloroform. 100 μl of blocking mixture (a mixture of sonicated human placenta DNA and cloned human repetitive sequences at a concentration of 1 mg/ml) was added, and the probe mixture was precipitated with ethanol. The nucleic acid was then resuspended in 20 μl of 5X saline sodium phosphate EDTA (SSPE), 0.1% sodium dodecyl sulfate (SDS), and prehybridized for 5 minutes at 42°C to saturate repetitive sequences which might be present in the probe. The probe was then added to a plastic bag containing a replica of the matrix filter and hybridization buffer (5X SSPE, 50% formamide, 0.2% SDS, 1X Denhardt's solution (i.e., 0.2% Ficoll, 0.2% polyvinyl pyrrolidone, 0.2% bovine serum albumin; see Denhardt, Biochem. Biophys. Res. Commun. 23:641 (1966)), and 20 μg/ml salmon sperm DNA) and the hybridization reaction carried out for 12 to 18 hours. Filters were washed in 0.1X SSPE, 0.1% SDS, at 65°C and exposed to X-ray film for 2 to 8 hours.

End-specific Nucleotide Sequence Analysis

Cosmids are transferred from archives to fresh 96-well plates containing liquid LB media with 25 μg/ml kanamycin sulfate and incubated at 37°C in a humidified atmosphere for 6 to 10 hours. DNA is prepared using a previously described cosmid miniprep procedure [Evans et al., Methods in Enzymology, Supra]. DNA from each cosmid is sequenced in a commercially available DNA Sequencer following the manufacturer's instructions.

Restriction Enzyme Analysis

Restriction enzyme analysis of isolated cosmids was carried out using DNA isolated from minilysates. Cosmid DNA was prepared from minilysates as follows:

DNA was isolated from 1.5 ml cultures. A culture was inoculated with a single bacterial colony and incubated with vigorous shaking at 37°C for 6 hours. DNA was prepared using a modified boiling procedure [Evans et al., Methods in Enzymology 152:604 (1987)]. Cells were collected by a brief (1 min.) centrifugation in a microcentrifuge and cells were resuspended in 300 μl of STET buffer (50 μM TRIS·HCl, pH 8.0, 8% sucrose, 5% Triton X-100 and 50 mM EDTA). 20 μl of freshly prepared lysozyme (10 mg/ml) in STET buffer were added, the mixture vortexed and incubated in a boiling water bath for one minute. The solution was immediately centrifuged for 10 minutes in a microcentrifuge and the gelatinous pellet removed with a toothpick and discarded. 325 μl of isopropanol were added and the mixture incubated at room temperature for 5 minutes. The precipitated DNA was collected by centrifugation at room temperature in a microcentrifuge, the pellet dried and resuspended in water.

DNA was digested to completion with *Not*I, digested partially with one or more enzymes (typically *Bam*HI, *Eco*RI, *Hin*dIII, *Sac*II, *Pvu*II, and *Kpn*I), separated on an agarose gel, transferred to a nitrocellulose filter and hybridized with ^{32}P-labeled oligonucleotides recognizing the T3 or T7 bacteriophage promoters. T3 and T7 oligonucleotides (commercially available as sequencing primers, Stratagene Cloning Systems, San Diego, Ca, USA) were labeled using polynucleotide kinase and γ-^{32}P ATP to a specific activity of 2 x 10^8 cpm/μg. The labeled oligonucleotides were then hybridized to the filters in 6X saline sodium citrate (SSC), 10% Denhardt's solution for 12 hours at 42°C and washed in 2X SSC for 10 minutes at 50°C. Filters were exposed to X-ray film for 20 minutes to 12 hours. The pattern of bands appearing on the autoradiograph could then be interpreted as indicating the distance from the cloning site to the restriction site, much as with the "cos"-mapping procedure of Rackwitz et al., Gene 30:195 (1984).

Alternatively, programmed automatic restriction enzyme digestions were carried out to completion in 96-well microtitre plates using a laboratory robot (Beckman Biomek 1000).

Data Analysis

The resulting hybridization data were manually entered into a computer file and analyzed using computer programs written in Turbo Pascal (Borland International) running on Apple Macintosh II or Macintosh SE computers. One program "Multiplex-mapper" compared data sets from hybridization reactions using mixed probes, determined those clones which were identified by more than one probe mixture, and produced a list of linked clones. A second program, "Contig-maker" assembled the list of overlapping clones into potential contigs which could be analyzed in

greater detail. In some cases, orientation and overlap of individual cosmid clones in a contig were confirmed by detailed restriction mapping and hybridization analysis of the individual cosmid clones.

Although data analysis was performed using the above-mentioned computer programs, a person of ordinary skill in the art should have no difficulty in carrying out the comparison of sequence and/or hybridization data and assembling the overlapping clones into contigs using other software. Moreover, manual data comparison and contig making are also possible, though more laborious.

EXAMPLE 2

Cosmid Multiplex Analysis of Human Chromosome 11q

The cosmid vector sCOS-1 (Figure 1) was used to prepare a genomic library from a somatic cell hybrid containing as its only human material DNA from the distal long arm of human chromosome 11, including 11q21-11qter, in a mouse background [Maslen et al., Genomics 2:66 (1988)]. The distal long arm of human chromosome 11 is of biological interest for a number of reasons. Like the major histocompatibility complex, the T cell receptor and immunoglobulin genes, and the IgK-CD8A-CD8B region of chromosome 2p12, human chromosome 11q23 contains a cluster of genes encoding proteins which are members of the immunoglobulin superfamily and are possibly important for cell-cell interactions in the immune and nervous systems including Thy-1, CD3, δ, epsilon, and N-CAM [Nguyen et al., J. Cell. Biol. 102:711 (1986)]. 11q23 is the location of genes in which defects may be responsible for ataxia telangiectasia [Gatti et al., Nature 336:577 (1988)] and other hereditary disorders including multiple endocrine neoplasia type I [Larsson et al., Nature 332:85 (1988)], diabetes analogous to the NOD mouse [Prochazka

et al., Science 237:286 (1987)] and others are also likely linked to markers on leukemias and pathognomonic for Ewing's sarcoma, peripheral neuroepithelioma and Askin's tumor [Griffin et al., Proc. Natl. Acad. Sci. USA 83:6122 (1986)]. The initial physical analysis of human chromosome 11 should allow eventual analysis of the genes associated with these phenomena and the underlying biology.

A genomic library consisting of 1.2×10^7 individual members was prepared and cosmids containing only human DNA were selected from this library by screening with probes recognizing human repetitive sequences. The proportion of human clones in this genomic library was 0.9%, indicating that the proportion of human chromosome 11 present in the somatic cell hybrid was about 27 mb, consistent with previous cytogenic and molecular characterization of this cell line [Maslen et al., Supra]. 1296 clones were selected, archived in 96-well microtitre plates, and arranged on a nitrocellulose filter according to the columns and rows of a 36 x 36 matrix. Using probes recognizing many available DNA markers mapping to this chromosome, consmids containing the genes THY1 [van Rijs et al., Proc. Natl. Acad. Sci. USA 2:5832 (1985)], T3D, T3E [Evans et al., Immunogen 28:365 (1988)], ETS1 [Watson et al., Proc. Natl. Acad. Sci. USA 83:1792 (1986)], PBG [Wang et al., Proc. Natl. Acad. Sci. USA 78:5734 (1981)], PGR [Misrahi et al., Biochem. Biophys. Res. Commun. 143:740 (1987)], SRPR [Lauffer et al., Nature 318:334 (1985)], and APOA1 [Karathanasis et al., Proc. Natl. Acad. Sci. USA 80:6147 (1983)]. The identified genes and clone coordinates for DNA markers on human chromosome 11q-11qter represented in the ordered cosmid set are shown in Table 2.

Table 2

Identified genes and clone coordinates for DNA markers on human chromosome 11q21 - 11qter represented in the ordered cosmid set.

cosmid clone coordinate (Y,X)	marker
3,11	THY1
11,34	THY1
10,7	T3D (CD3γ,δ)
7,13	ETS1
5,13	ETS1
4,22	PGR
13,27	APOA1
11,25	PBGD
12,19	D11S23
24,12	D11S24
24,8	SRPR
11,31	SRPR

Additional available RFLP markers [Maslen et al., Supra] were also identified in this collection to allow eventual correlation of the emerging physical map of chromosome 11 with the linkage map.

Groups of clones corresponding to 32 of the rows and 36 columns were pooled, and 68 hybridization reactions were carried out to replica filters according to the strategy outlined hereinbefore. Mixed probes detected a minimum of nine and a maximum of 46 cross-hybridizing unique clones on the filter matrix with each hybridization reaction using a pooled probe. When hybridization is carried out with a mixed probe consisting of RNA transcripts from cosmids of a row of the matrix, and a mixed probe representing a pool of all cosmids aligned along a column of the matrix, the cosmid clone which hybridizes with both mixed probes is linked to the clone located at the intersection of the row and column from which probe mixtures were prepared. To aid in the analysis of the data generated by this procedure, the Y and X coordinates of the cross-hybridizing clones are entered into a computer and matches identified using

one of several computer programs. From this series of experiments, 1099 linked clones were detected from the hybridization of 36 pooled columns and 32 pooled rows of the matrix. Several of these overlapping clones were analyzed by restriction mapping to confirm that the clones indeed did overlap in the expected manner.

Completeness of the Cosmid Multiplex Data

From the list of linked clones produced by this multiplex technique, contigs were assembled either manually or through computer analysis of the data from the predicted hybridization linkage using mixed multiple RNA probes. Based on an initial analysis of the data using a simple algorithm for contig construction, 315 contigs were assembled from the 1099 linked clones determined from multiplex analysis. The size of the contigs ranged from 2 linked cosmids to 27 cosmids grouped into a contig extending over several hundred kb, with the majority of contigs consisting of between 2 and 5 cosmids. To confirm that these groupings reflected the true structure of the human chromosome, and not artifactual groupings due to random cross-hybridization, several of the contigs were restriction mapped in detail to determine the degree of overlap and establish a physical map. The restriction map of a representative contig assembled by this strategy is shown in Figure 5.

Assessment of Progress

Based on the assumption that the region of human chromosome 11 carried by the parent hybrid represents about 27 mb, the collection of 1296 cosmid clones analyzed here represents about 2 genome equivalents. It is also estimated that the minimal detectable overlap by hybridization analysis using end-specific RNA probes is about 200 nucleotides. If θ is

the fraction of length of two clones which must be shared in order for overlap to be detected [Lander et al., Supra], then the expected number of contigs consisting of at least two clones generated by the analysis of N cosmid clones is

$$Ne^{-c(1-\theta)} - Ne^{-2c(1-\theta)}$$

wherein the redundancy of coverage, c = LN/G, where L is the length of the clone insert and G is the haploid genome length in base pairs.

The minimum detectable overlap with end-specific RNA probes is θ=0.005. Approaching the theoretical limit of θ=0, a maximum of about 450 contigs would be expected to result after the analysis of one genome equivalent and about 260 contigs after the analysis of 2 genome equivalents. Thus the analysis of the clone set carried out here, generating 315 contigs after the analysis of about two genome equivalents, is in good agreement with theoretical predictions. The main advantage of the current strategy is that the analysis of 1296 clones required only 72 analytical reactions, rather than 1296.

It was found that the prehybridization of the RNA probes with a high concentration of human repetitive sequences, as hereinabove described, was sufficient to completely block hybridization of most of these frequencies, and was sufficient for eliminating most of these artifactual linkages. However, the analysis of several large contigs mapping to human chromosome 11 generated by this analysis has revealed several cosmid clones which were included in a contig but which could not be substantiated based on the result of restriction mapping and hybridization analysis. This artifact may be the result of cryptic low-frequency repetitive or redundant sequences present in this region of the genome, or could be the result of genomic sequences which are

unstable and deleted or rearrange when cloned in *E. coli*. Evidence for the later sequences, isolated through screening non-amplified cosmid libraries, has been found in the analysis of the human CD3 locus [Evans et al., Immunogen, Supra]. However, it should be noted that the multiplex technique of the present invention, when carried to completion using both T3 and T7 mixed RNA probes, generates data that is internally redundant in that both members of a linked pair should cross-hybridize with one another. Thus, further refinement of this approach should eliminate most serious artifacts arising during multiplex clone analysis.

In this regard, the analysis disclosed in the present invention has generated a partially overlapping cosmid set which is estimated to include about 60% of the 11q21-11qter region of human chromosome 11q. The results of certain preliminary restriction enzyme analyses, further analysis of contigs and filling-in by traditional chromosome walking are in complete agreement with theoretical calculations of fingerprinting efficiency. A more complete analysis of this and other chromosome regions using a number of cosmids for 4 or more genome equivalents would be expected to result in near closure of the map. Using the technique of the present invention, this would require a collection of about 3600 cosmids and 120 T3 or T7 reactions/hybridizations rather than the 72 carried out in the present Example. In addition, the technique of the present invention is applicable for traditional chromosome "walking" to allow "filling-in" of gaps in a near complete map.

Additional analysis of this cosmid set representing chromosome 11q can be completed by automated restriction mapping. Analysis to date has revealed the presence of 177 potential "linking" clones, containing one or more *Not*I restriction sites, and 77 clones

containing *Sac*II sites indicative of hypomethylated CpG-rich islands. Forty of these cosmid clones contain clustered rare CpG rich restriction sites and can be identified unequivocally as hypomethylated islands. In addition, cosmid clones have recently proved very useful for *in situ* hybridization to metaphase or interphase chromosomes [Lichter et al., Proc. Natl. Acad. Sci. USA 85:9664 (1988)] and the identification of the cytogenic location of single-copy DNA sequences. These procedures potentially will allow ordering cosmid contigs with resolution of greater than 500 kb and, coupled with the strategy described here, provide a powerful mechanism for the constructions of physical maps of chromosomes.

Still further analysis of this cosmid set representing chromosome 11q can be carried out by direct automated DNA sequencing of the fragment ends as described above in the Detailed Description and in Examples 3 and 4.

EXAMPLE 3

Sequence-Sampled Map of *G. lamblia* genome

A 10.5 mb genome of *Giardia lamblia* (Fan et al., 1991, Nucleic Acids Res, 19:1905-1908) was cloned as a twenty-genome equivalent cosmid library. Five thousand cosmids can be mapped, end-oriented and end-sequenced generating 10000 ordered sequence fragments spaced, on average, every one kilobase. The determination of 500 bp of DNA sequence directly from each cosmid end results in an average spacing between islands of 0.5 kb.

Cosmid clones

A *Giardia lamblia* cosmid library was constructed in vector sCos-1. The library was prepared by partially digesting WB strain Giardia lamblia genomic

DNA with *Sau*3A and cloning the digestion products into the *Bam*HI site. From primary platings, 6,250 clones, representing about 20X coverage were picked and individually archived in 96-well microtitre plates and stored as frozen glycerol stocks. Clone names are derived from the library designation (GR) followed by the plate number and well position (e.g., cGR-2b11 denotes a cosmid located on plate 2 in well b11). These were arrayed onto filters with a Biomek 1000 robot (Beckman Instruments, Fullerton, CA) and the DNA fixed onto the filters using well known methods for analysis by hybridization. A specific hybridization probe to the β-giardin region was used to detect sixteen overlapping cosmids and a random probe detected another 26 overlapping cosmids. These cosmids were subjected to a series of mapping and sequencing experiments.

Automated DNA sample preparation

Template cosmid DNA was prepared by an alkaline lysis procedure (Sambrook et al., supra), or by using one of several DNA prep robots. Automated procedures used DNA prepared by an Autogen 540 DNA preparation robot (cycle 411), which was subsequently digested for one hour with RNAse A (75 μg/ul) in a total volume of 23 μl, and then precipitated with ethanol. Overnight growth of cosmid clones in 5 mls of terrific broth (see, Sambrook et al., supra) with 10 μg/ml kanamyacin generally provided sufficient DNA for one or two sequencing reactions.

Automated DNA sequencing

DNA sequencing was carried out using primers complementary to the T3 or T7 polymerase promoter located in the cosmid vector flanking the insertion site. Template cosmid DNA was prepared by an Autogen 540 DNA

preparation robot (cycle 411). Automated sequencing reactions were carried out using dye labeled T3 or T7 oligonucleotide primers with reactions assembled and cycle sequenced using the Applied Biosystems (ABI, Foster City, CA) Catalyst 800 robot with DNA concentrations of ~0.2 μg/μl. Sequence determination was carried out using the ABI 373A fluorescent sequencer. The labels (names) given to each of the DNA sequences determined from the cosmid clones are the clone names followed by -t or -u, denoting sequences from the T3 or T7 priming sites, respectively (e.g., cGR-19a9-u is the sequence of the T7 end of the cosmid clone found on plate 19 in position A9 in the GR library). The increased organization permitted by the new nomenclature system used in Examples 3 and 4 will facilitate the large scale physical mapping strategies described herein.

Contig construction and physical mapping

Identification of cosmids from the β-giardin genomic region was accomplished using filters representing five-fold of the twenty-fold genomic library. Bacterial clones were stamped into a grid pattern with a Biomek 1000 robot using S & S Nytran filters with a 0.4 μm pore size and fixed to the filters using standard fixation procedures. A probe of approximately 1 kb recognizing β-giardin genomic DNA was generated by PCR with the oligonucleotides GGTCAAGCTCAGCAACATGA (SEQ ID NO:1) and TGCTTTGTGACCATCGAGAG (SEQ ID NO:2) with standard amplification conditions and an annealing temperature of 60° C. Similarly, a random probe was generated as a 1.5 kb product with the primers CAGCAGATGGTCAAGCAAAA (SEQ ID NO:3) and ACTCCTGACACCACCACCTC (SEQ ID NO:4).

Physical and sequence maps were constructed by full digestion of each cosmid with the restriction

enzymes *Nco*I and *Bgl*II. Restriction fragments were separated on a 0.4% high strength agarose gel (0.5X TBE; see Sambrook et al., supra) run for 24 hours at 22v in a 20 by 20 cm gel apparatus. The *Nco*I and *Bgl*II restriction enzymes digest the vector opposite the cloning site generating genomic fragments. Each end of the genomic insert was detected as a vector/genomic chimera by hybridization with probes flanking the T3 and T7 promoter sites of sCos-1. The 1046 bp t3 probe was amplified from sCos-1 with the primers (5' to 3') TCGCTCACTGACTCGCTG (SEQ ID NO:5) and AGCCCTCCCGTATCGTAGTT (SEQ ID NO:6), and the 1004 bp T7 probe with the primers CTTGAGAGCCTTCAACCCAG (SEQ ID NO:7) and AACTGGGCGGAGTTAGGG (SEQ ID NO:8) with an annealing temperature of 60°C. and the standard conditions described above. The T7 probe was labeled by random priming with ^{35}S dATP and the t3 probe with ^{33}P dATP for dual-label hybridizations. Maps were constructed by determining an order of fragments with no gaps with the GRAM program [see, e.g., Soderlund et al., in *Proceedings of the 26th Hawaii International Conference on System Sciences: Biotechnology Computing*. (ed. Hunter, L.) 620-630 (CA: IEEE Computer Society Press, 1993)].

The mapping strategy consisted of digesting each cosmid with two different restriction enzymes and precisely determining the number and sizes of the various products. The enzymes *Nco*I and *Bgl*II were chosen for mapping because their recognition sites have different numbers of G/C bases and their sites are located several kb from the vector cloning site. Fragment sizes were estimated with the GelReader program of T. Redman (National Center for Supercomputer Applications, Urbana, Illinois) and maps constructed with a modified version of GRAM, which takes the end fragments into account. Maps of oriented cosmids were generated by comparison of the two maps of restriction fragments (Figure 6). The ends

were placed by fit relative to the neighboring sites for 32 of the 42 cosmids while the remaining ten were determined by the concordance of the possible orientations of ends on both enzyme maps using neighboring known cosmid ends as anchors. One 11h4 was equivocal and the FASTA analysis of the derived sequences showed that the t7 end matched the reverse complement of the 25h9 t3 sequence.

Sequence Analysis

The presence of repetitive sequences was determined using the program FASTA which compares all sequences to those previously determined supplemented with a comprehensive set of di- and tri-nucleotide repeats. A FASTA cutoff score of 100 was used to recognize repetitive sequences from background random matches. Similarities to known genes were identified with the BLAST program and the GenBank database. Amino acid comparisons were performed by translating DNA sequence fragments into all six potential reading frames and comparing translations to protein sequences in the non-redundant Swiss-Prot, GenPept or PIR database of the National Center for Biotechnology Information (Bethesda, MD) using the program BLASTX. The results of these various searches were evaluated numerically and by inspection (see, e.g., Table 3). The data obtained as described herein, including DNA sequence file pointers, matches from sequence analysis and information about overlapping clones, were stored in a relational database. The sequences generated by automated fluorescent sequencing from cosmid ends have been deposited with GenBank and were not edited to remove unidentified bases or correct the sequence.

Table 3 . Sequence searches performed with BLAST which produced likely, possible and interesting similarities to known sequences[a] with the sources of sequences, scores, percent identical residues and the length of amino acids matched. The highest scoring BLAST match is shown when more than one region or protein was similar to the unknown sequence.

sequence	similar to:	score	%ID	length
Likely				
cGR-38b8-t[a]	DNA similar to 3' noncoding region surface antigen gene, *G. lamblia* (gb:L11331)	761	100	154
cGR-45h5-u	protein kinase nek1, mouse (gb:S45828)	189	58	58
cGR-41e11-u	VSP1267 surface antigen, *G. lamblia* (sp:TSA4_GIALA)	189	58	60
cGR-62e12-u	protein kinase nek1, mouse (gb:S45828)	170	44	75
cGR-15c7-t[b]	G2 specific protein kinase, *A. nidulans* (sp:NIMA_ASPNI)	95	36	50
cGR-45h5-t	surface protein 11, *G. lamblia* (gb:M95814)	75	33	48
Possible				
cGR-16g10-u	transcription factor Oct-1, mouse (sp:HMO1_MOUSE)	74	34	58
cGR-15c7-u	MYB Protooncogene protein, *D. discoidium* (gb:S93742)	74	40	37

[a] Searches presented were protein sequence based with the exception of an identical nucleotide sequence found for cGR-38b8-L
[b] The overlapping sequence determined from cGR-15c7-t and cGR-6a6-u were searched together as one

Table 3. Sequence searches performed with BLAST which produced likely, possible and interesting similarities to known sequences[a] with the sources of sequences, scores, percent identical residues and the length of amino acids matched. The highest scoring BLAST match is shown when more than one region or protein was similar to the unknown sequence.

sequence	similar to:	score	%ID	length
Interesting				
cGR-51a9-u	eukaryotic initiation factor 4E, mouse (sp:IF4E_MOUSE)	70	32	62
cGR-19d7-t	polyposis coli protein DP1, human (gb:M73547)	69	38	42
cGR-57a10-u	ribonuclease inhibitor, rat (sp:RINI_RAT)	66	42	45
cGR-22b8-t	dynein associated (glued) protein, *D. melanogaster* (sp:DY15_DROME)	66	46	28
cGR-25h9-t	cosmid F09G8, *C. elegans* (gb:L11247)	66	27	51
cGR-2g8-u	ORF in thr4-sup61 intergenic region, *S. cerevisiae* (sp:YCV0_YEAST)	64	32	31
cGR-19g10-u	M5 protein precursor, *S. pyogenes* (sp:M5_STRPY)	63	29	54
cGR-16g8-u	zinc finger proteins, *X. laevis* (pir:S06546)	62	57	21
cGR-16g10-t	lamin B2, chicken (sp:LAM2_CHICK)	60	30	30
cGR-41f7-u	ankyrin 2, brain, human (pir:S14569)	57	33	54

[a] Searches presented were protein sequence based with the exception of an identical nucleotide sequence found for cGR-38b8-L
[b] The overlapping sequence determined from cGR-15c7-t and cGR-6a6-u were searched together as one

Sequences derived from cosmid ends were analyzed for the presence of repetitive sequences, simple sequence repeats and similarities to known genes (Figure 7). A number of genes were detected during the characterization of sequences as described herein. There were three protein kinases and a surface antigen gene in the random genetic region; no additional genes were detected near the β-giardin gene. Additional genes could likely be determined with gene prediction programs developed for these applications such as GRAIL.

Computer-aided primer pair design

One advantage of mapping and sequence sampling is future independence from clone libraries through PCR amplification. This advantage was demonstrated within the random region contig. The selection of primer pairs for PCR analysis was carried out using the PRIMER program provided by E. Lander (MIT) for each of the cosmid-derived end-sequences that were determined. Analysis was done in batch processing mode on a Sun workstation specifying an annealing temperature of 60°C and a primer length of 18 - 22 nt. Modifications of these parameters (oligonucleotide length 25 nt and annealing temperatures of 55°C for AT-rich sequences and annealing temperatures to 65°C or greater for GC-rich sequences) generally allowed production of a suitable primer. Primers were produced commercially by Genset, Inc. (Paris, France).

Predicted STS primers were tested by PCR amplification in a 30 l reaction volume containing: 10 mM Tris (pH 8.8), 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% gelatin, 200 M each dNTP, 100 ng genomic DNA and 1.5 units of Taq DNA polymerase. Initial PCR conditions were:

denaturation at 93°C for 2 minutes;

35 cycles of 30 seconds each at 94°C;

annealing for one minute at the predicted annealing temperature, and
30 seconds at 72°C; followed by
a final extension at 72°C for five minutes.

The probes were labeled by random priming with a ^{32}P dCTP and sCos-1 DNA was labeled with ^{35}S dATP to identify specific clones in the hybridization process.

The ordered sequences were then checked by PCR amplification demonstrating amplification of a specific region of interest independent of the cosmid libraries. Products were obtained for most (>85%) of the neighboring sequences. Some failures, possibly due to incorrect primer sequences and large distances between primers were not analyzed further within the random contig region.

In summary, two regions of the *Giardia lamblia* genome were mapped and sequences sampled from within those areas of approximately 160 kb (Figures 6A and 6B). The sequenced cosmid ends encompass approximately 15% of this total region. The average sequence read was 347 bp which does not differ from the median (353 bp) and has a standard deviation of 39 bp. Eight cases of cosmids ending at exactly the same *Sau*3A sites were observed and in four cases cosmid ends were close enough to one another to form small sequence contigs. Overall, the median spacing between cosmid ends was 1.25 kb with an average spacing of 2.0 kb, suggesting that the relatively rare events of cloning from defined restriction sites follows a poisson process suggested by visual inspection of Figure 8. These are underestimates of sequence overlaps and identical ends since cosmids which intrude into the contig without encompassing the region detected with the hybridization probe were not included in the contig. Full characterization of the region would nearly double the amount sequenced to approximately 30%. The large number of identical ends compared to overlapping

sequence contigs suggests that the practical limit for constructing a library with one restriction enzyme for partial digests is below twenty fold and perhaps as low as five to ten fold in the sequence sampling strategy.

EXAMPLE 4

Preparation of a Chromosome-11 Sequence Sampled Map Cosmid libraries

Two chromosome 11-specific cosmid libraries were constructed in vector sCos-1. Cosmids denoted 11q were isolated from a library prepared from hybrid TG5D1-1 representing chromosome 11q13-11qter (Evans and Lewis, 1989, Proc. Natl. Acad. Sci., USA, 86:5030-5034). TG5D1-1 is a Friend cell line derived from somatic cell hybrid 5D1 that carried an intact human X chromosome 11 [Pyati et al., Proc. Natl. Acad. Sci USA 77:3435 (1980)], and was selected for the loss of the entire X chromosome and most of chromosome 11. TG5D1-1 contains the distal portion of chromosome 11 as the only human material in a mouse genomic background [Maslen et al., Genomics 2:66 (1988)]. Cytogenetic and molecular analysis indicates that the amount of human DNA represented about 1% of the mouse genomic background [Maslen et al., Supra].

Cosmids denoted SRL were isolated from a flow sorted chromosome 11-specific library prepared from somatic cell hybrid J1 (described in Kao et al., 1977, Somatic. Cell. Genet., 3:421-429). The SRL library was prepared from 100 ng of flow purified chromosome 11 at the Los Alamos National Laboratory (L. Deaven and J. Longmire) as part of the National Gene Library Project and represented 125X coverage of chromosome 11. Approximately 17,000 clones, representing about 5X coverage, were picked from primary platings and individually archived in 96-well microtitre plates for use in this example. Characterization of this library

revealed that about 4% consist of non-human inserts. Clone names are derived from the library designation followed by the plate number and well position (e.g., c11q-2b11 denotes a cosmid from the c11q cosmid library located on plate 2 in well b11).

Cosmid end sequencing

To generate sequence data from the cosmids, DNA sequencing was carried out using primers complementary to the T3 or T7 polymerase promoter located in the cosmid vector flanking the insertion site. Template cosmid DNA was prepared by an alkaline lysis procedure (Sambrook et al., supra), or by using one of several DNA prep robots. Automated procedures used DNA prepared by an Autogen 540 DNA preparation robot (cycle 411), which was subsequently digested for one hour with RNAse A (75 μg/μl) in a total volume of 23 μl, and then precipitated with ethanol. Overnight growth of cosmid clones in 5 mls of terrific broth (Sambrook et al., supra) with 10 ug/ml kanamyacin generally provided sufficient DNA for one or two sequencing reactions. Alternatively, DNA template preparations were done using the custom "Prepper, Ph.D." DNA preparation robot developed by H. R. Garner (Garner et al., Sci. Computing and Automation, March/1993, 61-68). Some of the initial sequences in this example were determined manually using Sequenase kits with dideoxy terminators (U. S. Biochemicals).

Automated sequencing reactions were carried out using dye labeled T3 or T7 oligonucleotide primers with reactions assembled and cycle sequenced using the Applied Biosystems (ABI) Catalyst 800 robot with DNA concentrations of ~0.2 μg/μl. Automated DNA sequencing was carried out using the ABI 373A sequencer. The names of DNA sequences determined from the cosmid clones are the clone names followed by -t or -u, denoting sequences

from the t3 or t7 priming sites, respectively. (e.g., c11q-2b11-u is the sequence of the t7 end of cosmid 2b11 from library 11q).

A sample of 371 DNA sequence fragments were determined by automated sequencing from cosmid ends. Forty-nine sequences were not used for STS production in this example and were placed in reserve should additional STSs be needed. Extensive regions of repetitive sequence which made PCR primer prediction unlikely occurred in 14% (45/322) of those analyzed (Table 4). The remaining 277 sequences (86%) were used for STS primer prediction by computer analysis and oligonucleotide primers were synthesized and tested.

TABLE 4

Analysis of 322 cosmid-derived sequences for STS production

Fate of sequences	number	percentage of total
Not suitable for primer prediction	45	14%
failed STS production	65	20%
successful STSs production	212	66%
STSs mapped to a region	187	58%

Note: Cosmid sequences were determined by direct priming of cosmid templates using oligonucleotides complementary to the T3 or T7 polymerase promoters present in the vector. STS primers were predicted using the PRIMER program.

The sequences generated by automated fluorescent sequencing from cosmid ends have been deposited with GenBank and were not edited to remove unidentified bases or correct the sequence. The sequence information was not corrected before analysis so that the Primer program would not predict primers in regions of questionable accuracy. Consequently, some of the alignments showing similarity between known protein sequences and translated chromosome 11 specific sequences contain Xs at unknown and stop codons which could be due to errors in the sequence or translating beyond exons into introns.

Computer-aided primer pair design

The selection of primer pairs for STS analysis was carried out using the Primer program available from E. Lander (MIT). Analysis was done in batch processing mode on a Sun workstation specifying an annealing temperature of 60° C and a primer length of 18 - 22 nt. Modifications of these parameters (oligonucleotide length 25 nt and annealing temperatures 55°C for AT-rich sequences and annealing temperatures to 65°C or greater for GC-rich sequences) generally allowed production of a suitable primer set. DNA sequences that contained extensive regions of repetitive sequence so that primers could not be designed to generate products of at least 120 bp were not utilized. Primers were synthesized using an ABI PCR-Mate oligonucleotide synthesizer or produced commercially by Genset, Inc. (Paris, France).

Of this collection of primers, 77% (212/277) generated PCR amplification products from human genomic DNA and J1 hybrid DNA without producing products from yeast genomic DNA and are suitable for STS content mapping using YAC libraries. Localization of 88% of these new STS markers (188/212) was carried out by hybrid

analysis, FISH or a combination of both. For comparison, nearly the same success rate occurred in predicting primer sets from sequences retrieved from GenBank. In some cases primer sets derived from the cSRL library which failed to amplify human DNA clearly generated a product from hamster -- the J1 cell line host (Table 5).

STS MAP OF CHROMOSOME 11

Table 5

Hybrid cell lines containing specific regions of human chromosome 11

	chromosome 11 region			
Cell line	cytogenetic	flpter	species	reference
J1	p15.5-q25	0-1	hamster	Kao, et al., 1977
CH66-36	p15.5-q24.1	0-0.93	hamster	G. Hermanson and M. Altherr (unpub. data)
ALE4	p15.5-q23.3	0-0.85	hamster	Delattre, et al., 1991
MC-1	p15.5-q23.3	0-0.85	hamster	Glaser, et al., 1989
MX-11	p15.5-q23.3	0-0.85	mouse	Junien, et al., 1992
CY.8	p15.5-q14.1	0-0.6	mouse	D. Callen (pers. comm.)

Cell line	chromosome 11 region: cytogenetic	.flpter	species	reference
16TX2	p15.5-p11.2	0-0.33	mouse	Junien, et al., 1992
CH66-23	q24.1-q25	0.93-1	hamster	G. Hermanson and M. Altherr (unpub. data)
R28/4D	q13.1-q25	0.5-1	hamster	Glaser, et al., 1989
CF52-46	q13.1-q25	0.5-1	mouse	Junien, et al., 1992
POR4	p13.1-q25	0.24-1	mouse	Junien, et al., 1992
PEL 16	p15.4-q25	0.05-1	mouse	Junien, et al., 1992

Note: The species from which cell lines were derived and a reference is indicated above.

STS markers generated from mapped cosmids were supplemented with STSs corresponding to almost all of the cloned genes on chromosome 11 to allow for eventual production of a complete chromosome map. Genes mapped to chromosome 11 were identified by searching the Genome Data Base (GDB; Pearson et al, 1992, Nucleic Acids Res., 20:2201-2206), On-line Mendelian Inheritance in Man (OMIM), GenBank, Medline, and the like. The DNA sequences were retrieved from GenBank when available and STS primers designed using PRIMER. Whenever possible, chromosome 11 genomic DNA sequence was used for STS production and primer pairs were designed to generate products of 400 to 1000 base pairs. When genomic DNA sequence was not available, the generally intron-free 3' untranslated region of chromosome 11 cDNAs was chosen. Where these were not available, cDNA coding regions were utilized and the predicted amplification product sizes were limited to 150-250 bp in order to minimize the chances of including an intron in the target region. In the overall design of STS primer sets, predicted product lengths were intentionally varied over the range 150 to 1000 bp to allow for eventual multiplex screening of YAC libraries.

Characterization of chromosome 11-specific STSs

Predicted STS primers were tested by PCR amplification (Saiki et al., 1988, Science, 239:487-491) in a 30 μl reaction volume containing: 10 mM Tris (pH 8.8), 50 mM KCl, 1.5 mM MgCl2, 0.001% gelatin, 200 μM each dNTP, 100 ng template DNA and 1.5 units of Taq DNA polymerase. Initial PCR conditions were:

denaturation at 93°C for 2 minutes;
35 cycles of 30 seconds each at 94°C;
annealing for one minute at the predicted annealing temperature, and
30 seconds at 72°C; followed by

a final extension at 72°C for five minutes. Initially, three different concentrations of primers (833, 500 and 250 nM) were tested for amplification using human genomic DNA. An optimal concentration and temperature were determined for the amplification of the following test DNA samples: human genomic DNA, DNA from the hybrid cell line J1 containing human chromosome 11 in a CHO cell background, hamster genomic DNA, mouse genomic DNA and yeast DNA. Those primer sets that produced an amplification product of the expected size from human genomic DNA and J1 DNA were characterized further. Most of the STSs generated have a narrow range of annealing temperatures from 56°C to 60°C and relatively short primer lengths of around 20 nucleotides.

Physical Mapping by fluorescence *in situ* hybridization (FISH)

To map the genomic location of each STS, *in situ* hybridization using cosmid DNA was carried out on metaphases prepared from human fibroblasts (CRL1634; Human Genetic Mutant Cell Repository, Camden, New Jersey) using well-known methods (see, e.g., Giovannini et al., 1993, Cytogenet. Cell. Genet., 63:62). Chromosomal localization was determined as chromosome 11 FLpter (fractional chromosomal length from 11pter) and cytogenetic band position was extrapolated from the coincidence of FLpter values to cytogenetic bands on the chromosome ideogram.

Somatic cell hybrid analysis

Additional localization of some cosmid sequences, and localization of STSs produced from gene sequences, were carried out using a panel of hybrid cell lines set forth in Table 5 containing part or all of human chromosome 11 in mouse or hamster genomic

backgrounds. Analyses of the PCR amplification of STSs from DNA isolated from this set of hybrids mapped the STSs to eight distinct regions, or bins, on chromosome 11 (Figure 10).

The whole set of 370 STS markers, associated with known genes or chromosome-specific cosmid clones, was regionally mapped to chromosome 11. Some of the map locations of these cosmids had been reported previously, though not associated with STS identifiers of the present invention. Of the 370 STSs generated, 335 were successfully mapped to specific regions of human chromosome 11 (Table 6) using FISH or through the PCR analysis of a somatic cell mapping panel. Mapping information for some markers was available from GDB and was confirmed employing the techniques described herein based upon hybrid analysis. Consistent results were found in all cases when mapping information on many markers was determined using both FISH and hybrid analysis.

In general, the set of standard STSs derived in this example appear to be uniformly distributed throughout the chromosome and include markers for most of the significant mapping landmarks. Chromosome 11 was divided into twenty regions and the average distribution of STSs along chromosome 11 was assessed. The average number of STSs per region was 17 with a standard deviation of 4.5. The most over-represented region was FLpter 0.85 to 1 (11q23.3 - q25) while 0.3 to 0.45 (11p12 - q12.1) and 0.6 to 0.85 (11q14.1 - q23.2) contained proportionally fewer STSs. The relative precision of the mapping information can be estimated from the distribution of FLpter ranges of the 335 mapped STSs. Twenty-four percent of the markers had an FLpter range of less than 0.05, 55% less than 0.1 and 83% less than 0.2.

Table 6

Chromosome 11 specific STSs, mapping positions, technique[a], product size, primer names, sequences, concentrations and annealing temperatures

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S1002	c11q-1a7	0.50-0.60 11q13.1-q14.1	b	176	c11q-1a7-tA c11q-1a7-tZ	TCTGTCTGGAAGAGCGTGG CCAGACCAAGTCTTGGACCT	250 250	56
D11S493	c11q-1d7	0.56-0.64 11q13.5-q14.2	c	194	c11q-1d7-tA c11q-1d7-tZ	CCATAAATGAAAGGTTTGGATG GCGACTTGAGCTGTCATGC	250 250	56
D11S360	c11q-1f6	0.62-0.66 11q14.1-q14.3	a	252	c11q-1f6-tA c11q-1f6-tZ	TGGCATTATAGGTGCCCG CTCAGGAAACAGGGAGATGC	250 250	56
D11S366	c11q-2b11	0.86-0.89 11q23.3	ab	183	c11q-2b11-tA c11q-2b11-tZ	CCAGGAAGGGTTCCCTGAAGTAGGTG ATAAGCTGCTGCTACCTACAGCTGCC	250 250	56
D11S361	c11q-2c4	0.70-0.72 11q21-q22.1	ab	189	c11q-2c4-tA c11q-2c4-tZ	AGAAAACTGATTAACCAAGGCA AGCAAATCTCGTCTCAAAAAAA	833 833	56
D11S353	c11q-2d4	0.55-0.58 11q13.3-q13.5	ab	150	c11q-2d4-uB c11q-2d4-uX	TCCAGTAGGAAGAGATGCAGG GGATTCCTTCTACTCTCCCAGG	250 250	56
D11S356	c11q-2g9	0.56-0.59 11q13.4-q13.5	abc	183	c11q-2g9-tA c11q-2g9-tZ	AGCTGTGGGTTCATAATGGC AGGTGAAGAGGAGGAGAGTGC	833 833	62
D11S491	c11q-3c3	0.56-0.59 11q13.5	b	138	c11q-3c3-uA c11q-3c3-uZ	CCTGTGGCCTCCTATGACAT TGAGGCCTCTCTTCTCTCATG	833 833	56
D11S794	c11q-3f11	0.92-0.92 11q24.1	a	154	c11q-3f11-uA c11q-3f11-uZ	CTTGGGAGTCACTGAACTGATG CCCACATCCCTAAACTGGAA	250 250	56
D11S1004	c11q-3h11	0.85-0.85 11q23.3	b	170	c11q-3h11-uA c11q-3h11-uZ	AGGAAACTGAGGAACAGAGCA GGCTGGCCTGGAGTTAACTA	833 833	55

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S528	c11q-4e1	0.85-0.91 11q23.3	bc	177	c11q-4e1-uS c11q-4e1-uA	CTAGCATGTGAGGGTGTTGGAGGT AATCACACTCTTGGTGGGTTTTCG	440 440	55
D11S1005	c11q-5c3	0.85-0.93 11q23.3-q24.1	b	101	c11q-5c3-tB c11q-5c3-tY	AACACCAACATCACAAAAGAAA ACTGACAAGAGCTTGCTGCA	250 250	58
D11S795	c11q-5e5	0.92-0.92 11q24.1	a	216	c11q-5e5-tA c11q-5e5-tZ	AGCACACCCGGTGGTATG AGATGATGCCCTTCCCCC	500 500	51
D11S793	c11q-5l9	0.85-0.83 11q23.3-q23.2	bc	256	c11q-5l9-tA c11q-5l9-tZ	GGTCCCCTCTAAGAAGCACC CTTTTCCAAGTGCCAGCCT	250 250	56
D11S368	c11q-6c10	0.86-0.89 11q23.3	a	257	D11S368-A c11q-6c10-uZ	TTATAGGCGCCTGCACCACGTCT TCAAGAGGTTGAAGCAGTGAGCCGTG	500 500	51
D11S1006	c11q-6h6	0.93-1.00 11q24.1-q25	b	228	c11q-6h6-1A c11q-6h6-1Z	CACCATGCAATCTGCAGC ATTTCTGCCTGCTGATCCAT	250 250	58
D11S800	c11q-7d4	0.59-0.68 11q14.1-q14.3	b	150	c11q-7d4-uA c11q-7d4-uZ	TAGGAACCCCCATCATTCCA CCATGGAGGTCCCCTTTC	833 833	52
D11S492	c11q-7h10	0.78-0.85 11q23.1-q23.3	b	161	c11q-7h10-tA c11q-7h10-tZ	AAGCTCTCTGGGAAATAAAGGG AACAAACTCATGAATGCCTGG	833 833	56
D11S797	c11q-7h8	0.85-0.91 11q23.3	b	207	c11q-7h8-tA c11q-7h8-tZ	GTGTGGGGCCACTGTATTG TGGCAACAAAAGCAAGACTG	833 833	56
D11S488	c11q-8d1	0.91-1.00 11q24.1-q25	a	190	c11q-8d1-tA c11q-8d1-tZ	AAGGAAGGAAGGAAGGAAAGG CTGATAGCCTGACCTGACTGTG	500 500	58

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S783	c11q-8d10	0.94-1.00 11q25	a	230	marco2s marco2a	GGAGAGGACATTAACCAGCTGA GTGATGGGTTGCATAGTGTCTG	877 877	50
D11S784	c11q-8d11	0.94-1.00 11q25	a	154	c11q-8d11-lA c11q-8d11-lZ	GAGAGTGGAAGCAAAGGCTG CTCCCTTCCTTACCCTACAACC	250 250	58
D11S1008	c11q-8h4	0.60-0.85 11q14.1-q23.3	b	265	c11q-8h4-lA c11q-8h4-lZ	ACTGGGGACCCCATGACT GGGCTCCTGCTCTACCTCTT	500 500	56
D11S1009	c11q-9a1	0.94-1.00 11q25	b	250	c11q-9a1-lA c11q-9a1-lZ	ACTGTGTGGGGAGAAGGTACC TGCTGGTAACTTACCACTGTGC	500 500	56
D11S1001	c11q-10d7	0.85-0.93 11q23.3-q24.1	b	244	c11q-10d7-lA c11q-10d7-lZ	CCACCATATCCTGCCAAATT GGTTGCAAGATCACGTCACT	250 250	65
D11S1011	cSRL-1a2	0.94-1.00 11q25	a	150	cSRL-1a2-uA cSRL-1a2-uZ	CAGCTTTACTTTTATTTACAGAGGTTT CTAGAACTTCCTGGTTTAAACGAAA	833 833	56
D11S1010	cSRL-1a12	0.46-0.48 11q12.3	ab	180	cSRL-1a12-uA cSRL-1a12-uZ	GCTTGCTGCAATGTTTTTCA CTGGCTGCTAAAAGAAGAGACC	833 833	56
D11S1013	cSRL-1b2	0.95-0.98 11q25	ab	175	cSRL-1b2-uA cSRL-1b2-uZ	GAGTTTGGAGCAAGCCAGTC TGACCTTGATGCTCACTTGG	500 500	56
D11S1014	cSRL-1b4	0.61-0.61 11q14.1	ab	154	cSRL-1b4-uA cSRL-1b4-uZ	AAGTCACACAGCTAGCAAATGG TGTCCTTGACGGAATAGATTCC	250 250	58
D11S1015	cSRL-1b8	0.44-0.59 11q12.1-q13.5	a	201	cSRL-1b8-uA cSRL-1b8-uZ	GGTCAGCCCACACCTCAC ATACCCTTCAGGTGGATCCC	250 250	60

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S1012	cSRL-1b11	0.40-0.55 11p11.11-q13.3	a	151	cSRL-1b11-uA cSRL-1b11-uZ	GCTGTCAGACTGAAGACCCTG ACTGCCAGCCTGGTCAGAG	833 833	58
D11S1016	cSRL-1c5	0.79-0.85 11q23.1-q23.3	ab	185	cSRL-1c5-lA cSRL-1c5-lZ	TTCAGCACTCCTTACATTGCC TCCAACTGAAACAGCCTGC	833 833	58
D11S1017	cSRL-1c8	0.93-0.93 11q24.1	ab	175	cSRL-1c8-lA cSRL-1c8-lZ	GCACTCCTGTCTTGGTCCAT AGGGATCCATCTTGGAAGCT	250 250	58
D11S1018	cSRL-1d4	0.24-0.33 11p13-p11.2	b	211	cSRL-1d4-uA cSRL-1d4-uZ	CCAAGGACAGGCAGTAGAGC CAATGAAATGATGCAGTGTGG	250 250	58
D11S1019	cSRL-1d5	0.05-0.24 11p15.4-p13	b	141	cSRL-1d5-uA cSRL-1d5-uZ	CCTCAGTCCTTTGCACTTG TCTGGCTATTTTGTACAGGG	250 250	56
D11S1020	cSRL-1d6	0.05-0.12 11p15.4-p15.1	ab	241	cSRL-1d6-lA cSRL-1d6-lZ	GCCCTCTAGTCATGCCACAT ACGCGAAACCAGATCATTCT	500 500	58
D11S1022	cSRL-1l5	0.33-0.50 11p11.2-q13.1	b	134	cSRL-1l5-uB cSRL-1l5-uZ	CAACCAAAATAAAAGGCCTCC GTGTTCTTCTCTTGCACCTGC	250 250	56
D11S1023	cSRL-1l8	0.85-0.93 11q23.3-q24.1	b	215	cSRL-1l8-uA cSRL-1l8-uZ	GCAGGTTTTTCAAGAAGGTATG GGCAGCCATTTTCTTACCAA	833 833	58
D11S1021	cSRL-1l12	0.00-0.05 11p15.5-p15.4	b	157	cSRL-1l12-uA cSRL-1l12-uZ	CATGCAGGTGTCAGGACG GATACCAGAGAGATTGCATCCC	833 833	57
D11S1024	cSRL-1g5	nd		219	cSRL-1g5-uB cSRL-1g5-uY	GATGCCATGCTGGGTTAAAC TCATCTTAAGCCCATTCTCTCC	833 833	52
D11S1025	cSRL-1g8	nd		155	cSRL-1g8-uB cSRL-1g8-uY	CGACCCTGGGTGTTAACAAC AAAGGAACGTGTTACCGCTG	833 833	54

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	(primer) (nM)	Annealing temp.(°C)
D11S1026	cSRL-1g9	0.60-0.85 11q14.1-q23.3	b	151	cSRL-1g9-uA cSRL-1g9-uZ	GTTGATGACCCAGACAAAACTG TATCCAGTCATGTTACGTTTGC	833 833	56
D11S1027	cSRL-1h6	0.50-0.85 11q13.1-q23.3	b	172	cSRL-1h6-uA cSRL-1h6-uZ	GTCCAATACCGTACCCACTAGC CGATCTAGAGCTGTGCTGTCC	500 500	58
D11S1028	cSRL-1unk	0.50-0.85 11q13.1-q23.3	b	155	cSRL-1h10-uA cSRL-1h10-uZ	CGCAAGACTAACAAGCCCTC CAAACATTGCATATCCTGATGC	833 833	58
D11S1029	cSRL-2a2	0.85-0.93 11q23.3-q24.1	b	282	cSRL-2a2-tA cSRL-2a2-tZ	GACACACCATGAAGCACACC AGAGCCAATCTGCAAACACTG	250 250	56
D11S1030	cSRL-2b3	0.60-0.85 11q14.1-q23.3	b	177	cSRL-2b3-tA cSRL-2b3-tZ	CGAGACTCCATCTCACAAAAA TCTATCTGTGGTGCCCCAG	833 833	56
D11S1031	cSRL-2c1	0.05-0.24 11p15.4-p13	b	201	cSRL-2c1-tA cSRL-2c1-tZ	GCCCTATGTACTTTCTGTTGCC ACAGGAAGCTCAGGGCACTA	500 500	58
D11S1033	cSRL-2c2	0.33-0.50 11p11.2-q13.1	b	211	cSRL-2c2-tA cSRL-2c2-tZ	AAGCAACTTCCACATTCAAGTG GTGCAATGTCCCCTCATTTT	250 250	56
D11S1034	cSRL-2c3	0.05-0.24 11p15.4-p13	b	261	cSRL-2c3-tB cSRL-2c3-tY	AGGCTTTTCTAAACCTTCTCCC ACGGACGCAGATTCCTTTG	500 500	56
D11S1035	cSRL-2c4	0.93-1.00 11q24.1-q25	b	151	cSRL-2c4-tB cSRL-2c4-tY	CAGTAACTCCCAGGGCTCC CACTCACTGTAAATGCAGCCA	250 250	56
D11S1036	cSRL-2c5	0.60-0.85 11q14.1-q23.3	b	292	cSRL-2c5-tB cSRL-2c5-tY	CCCACAAATACCAAGATGACTG ACATCAATTTTGGCTGAGCC	250 250	56

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	(primer) (nM)	Annealing temp.(°C)
D11S1037	cSRL-2c7	0.93-1.00 11q24.1-q25	b	251	cSRL-2c7-1A cSRL-2c7-1Z	TCTTCGAGTCAGCAATTATCAA TACCTCCTTTTCTCCTACAAAG	250 250	56
D11S1038	cSRL-2c8	0.50-0.85 11q13.1-q23.3	b	205	cSRL-2c8-1A cSRL-2c8-1Z	ACAGACCACAGTCCCTCCAA CCCACTCCCCAAATACTCAA	250 250	50
D11S1039	cSRL-2c9	0.50-0.60 11q13.1-q14.1	b	210	cSRL-2c9-1A cSRL-2c9-1Z	TTCTCACCTTCACCAACCAC CTAAAATAACAAAACCCCCCC	250 250	56
D11S1032	cSRL-2c10	0.05-0.24 11p15.4-p13	b	172	cSRL-2c10-1A cSRL-2c10-1Z	CTCCACCCAACCTCACTCAC TCTACCCTACCCCACTTTCTCC	250 250	56
D11S1040	cSRL-2d1	0.60-0.85 11q14.1-q23.3	b	252	cSRL-2d1-1A cSRL-2d1-1Z	AACAATACACCATATTCTCCCC TACTTCACCTCAATCTTTCCCA	250 250	56
D11S1041	cSRL-2d3	0.85-0.93 11q23.3-q24.1	b	291	cSRL-2d3-1A cSRL-2d3-1Z	TCTCTCACCCACACCAATTC ACTTTCAACATCACATTCCACC	250 250	56
D11S1042	cSRL-2d7	nd		278	cSRL-2d7-1A cSRL-2d7-1Z	TCCCCTCCACACCACAAC AACCATCCCCTCACAATCAC	250 250	50
D11S1043	cSRL-2d8	0.05-0.24 11p15.4-p13	b	164	cSRL-2d8-1A cSRL-2d8-1Z	TTCTACACCTTTTTCTCCCTCCCTA CTTAACTAAAATTTCACCATACAAATT	833 833	49
D11S1044	cSRL-2e1	0.00-0.05 11p15.5-p15.4	b	276	cSRL-2e1-1A cSRL-2e1-1Z	CCCCACAAACTACACCCACA AACTCTCACCCCACACACTCA	250 250	56
D11S1048	cSRL-2e2	0.93-1.00 11q24.1-q25	b	203	cSRL-2e2-1A cSRL-2e2-1Z	TCTAACCTCCCTCCCCATAAC ATCTCTCTCACCACACCCTC	250 250	58
D11S1049	cSRL-2e3	0.05-0.24 11p15.4-p13	b	253	cSRL-2e3-1A cSRL-2e3-1Z	TCCTTCCTTTTTCTCACCATTC ATATTCTCATCCACCTCCCA	250 250	56

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S1050	cSRL-2e4	0.05-0.24 11p15.4-p13	b	281	cSRL-2e4-1A cSRL-2e4-1Z	CGGTTTCTCAGGCATGTGAT AATGAGGGTACACATCCACACA	250 250	56
D11S1051	cSRL-2e6	0.24-0.33 11p13-p11.2	b	252	cSRL-2e6-1A cSRL-2e6-1Z	ATACCAAAGCCCCTAAAACATG GTATCATATTTGCTGCTTCCCT	833 833	56
D11S1052	cSRL-2e7	0.33-0.50 11p11.2-q13.1	b	250	cSRL-2e7-1A cSRL-2e7-1Z	ACCCGACTCCCTTTCCAC TTGTCTACAAATCCCCCTCC	250 250	56
D11S1053	cSRL-2e9	nd		171	cSRL-2e9-1A cSRL-2e9-1Z	TGTCCATGCCTATGTCCTCA GGTGTTGGGAAAACTGGCTA	250 250	56
D11S1045	cSRL-2e10	0.93-1.00 11q24.1-q25	b	256	cSRL-2e10-1A cSRL-2e10-1Z	ACAGTTGATTATCCCATCAGGG AGCATGCACATGTCTTTGTTGG	250 250	56
D11S1046	cSRL-2e11	0.60-0.85 11q14.1-q23.3	b	176	cSRL-2e11-1A cSRL-2e11-1Z	GTCTCTGGCACACATTCCTCA TGCACATCTGTCCTGCTATTT	250 250	53
D11S1047	cSRL-2e12	0.05-0.24 11p15.4-p13	b	254	cSRL-2e12-1A cSRL-2e12-1Z	GGAAGACAAATCCCAATGCA GAGCCCAAGCGTAGGAAAC	500 500	50
D11S1056	cSRL-2i2	0.60-0.85 11q14.1-q23.3	b	203	cSRL-2i2-1A cSRL-2i2-1Z	ATGAGTCCCGTCATACTGTGC CCTTGTGGGAAAAAAGGCAG	250 250	56
D11S1057	cSRL-2i3	0.24-0.33 11p13-p11.2	b	172	cSRL-2i3-1A cSRL-2i3-1Z	CAAGTGTGTTCTGCCTTCACA CACAAGGGCACAGCTAGACA	833 833	56
D11S1058	cSRL-2i4	0.50-0.60 11q13.1-q14.1	b	260	cSRL-2i4-1A cSRL-2i4-1Z	TCAGCTGTATAAATGTCTGCCG GGTCAGAATGCCTGCACTGT	250 250	56

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S1059	cSRL-2f5	0.05-0.24 11p15.4-p13	b	256	cSRL-2f5-1A	TATTTGTCTTTCCATGCCTGG	250	56
					cSRL-2f5-1Z	TTTACTAAGCCTGAGAGCCAGC	250	
D11S1054	cSRL-2f10	nd		253	cSRL-2f10-1A	CTAACACAGGAACAGCAAGCC	250	56
					cSRL-2f10-1Z	ATGTCCAGAGTGTGCAGGTT	250	
D11S1055	cSRL-2f11	nd		267	cSRL-2f11-1A	AGCTTTTTTTGATGTTTGCTGG	250	56
					cSRL-2f11-1Z	GCAACTGCAACAAAAGCAAA	250	
D11S1061	cSRL-2g4	0.93-1.00 11q24.1-q25	b	267	cSRL-2g4-1A	ACAGTGCCCTGCCAAAGATG	250	56
					cSRL-2g4-1Z	AATTTCAGGTTTTTGTTAGGGC	250	
D11S1062	cSRL-2g6	0.05-0.24 11p15.4-p13	b	224	cSRL-2g6-1A	GCGTGAAGGCTGGATTCTAG	250	56
					cSRL-2g6-1Z	AACAAGAGTTTCCTCCAAGGC	250	
D11S1063	cSRL-2g7	0.60-0.85 11q14.1-q23.3	b	151	cSRL-2g7-1A	ATGAGCATGTTTTTCTGCTACA	833	56
					cSRL-2g7-1Z	AGCCAACCAACAGTTAAATTGA	833	
D11S1064	cSRL-2g9	0.05-0.24 11p15.4-p13	b	154	cSRL-2g9-1A	TGCACGTGAGATGAGTCTCC	250	56
					cSRL-2g9-1Z	CCACTGCAAAAACACGCTAA	250	
D11S1060	cSRL-2g11	0.05-0.24 11p15.4-p13	b	198	cSRL-2g11-1A	TTTGTTGTTTAAGTCACCCCG	250	56
					cSRL-2g11-1Z	CCCCTGATCTCAGACTTCCA	250	
D11S1067	cSRL-2h2	0.50-0.85 11q13.1-q23.3	b	264	cSRL-2h2-1A	GAGAACCCCTGTGTTACCTCC	250	56
					cSRL-2h2-1Z	AAAACCACAATGCCATCACA	250	
D11S1068	cSRL-2h9	0.05-0.24 11p15.4-p13	b	204	cSRL-2h9-1A	AGCCTCCCTCATCATTTGG	250	56
					cSRL-2h9-1Z	ACTGCTTTTGCACATGTACCC	250	
D11S1065	cSRL-2h10	0.60-0.85 11q14.1-q23.3	b	302	cSRL-2h10-1A	AGAGGTACCCTGCTTAGGATTC	500	56
					cSRL-2h10-1Z	TCATTCCTCACTCTCAATGCA	500	

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FISH range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S1066	cSRL-2h11	0.33-0.50 11p11.2-q13.1	b	260	cSRL-2h11-1A cSRL-2h11-1Z	TGTCGGTCCAGTTTTC CCCTAGCAACGACACACA	250 250	56
D11S1069	cSRL-3a1	0.33-0.50 11p11.2-q13.1	b	171	cSRL-3a1-1A cSRL-3a1-1Z	CTTTTTATTTCCCCAAATGGC CTGAGGCCTCCTTGTACTGC	500 500	56
D11S1070	cSRL-3a3	0.24-0.33 11p13-p11.2	b	207	cSRL-3a3-1A cSRL-3a3-1Z	GCTAATATTTTGAGTCAATTTTCCTC CAAAACTATGCTAAAGCCTGTATATG	833 833	56
D11S1071	cSRL-3a4	0.03-1.00 11q24.1-q25	b	275	cSRL-3a4-1A cSRL-3a4-1Z	GGTCCCTCTGACCTCACATT ATCCTACCCACCCTACCCAT	250 250	56
D11S1072	cSRL-3a6	nd		172	cSRL-3a6-1A cSRL-3a6-1Z	GGCATCTTGCCCTGTTTC CCCTGGGAACAAAACAAA	250 250	56
D11S1073	cSRL-3a7	0.60-0.85 11q14.1-q23.3	b	158	cSRL-3a7-1A cSRL-3a7-1Z	CAGGAGCCAGAATGCAACAC ATTCCTACTGTCGGGCCC	250 250	56
D11S1074	cSRL-3a8	0.50-0.85 11q13.1-q23.3	b	169	cSRL-3a8-1A cSRL-3a8-1Z	CAGATGAAAAGATCTGCACCA CATCCCTCCTGACCACACTT	250 250	56
D11S1075	cSRL-3a9	0.05-0.24 11p15.4-p13	b	227	cSRL-3a9-1A cSRL-3a9-1Z	AGCACCTGAACCCTTCC CAATAGCACTGTTTTGAAGGGG	250 250	56
D11S1076	cSRL-3b1	0.33-0.50 11p11.2-q13.1	b	155	cSRL-3b1-1A cSRL-3b1-1Z	AGACTGGTTGGGCAAGGC GAGTCCTGTCCTCCTTGCTG	250 250	57
D11S1080	cSRL-3b2	0.33-0.50 11p11.2-q13.1	b	317	cSRL-3b2-1A cSRL-3b2-1Z	CACCATGGGTATCCAGTGC ACCAGCCCCTAGAATTCCAG	500 500	56

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S1081	cSRL-3b3	0.34-0.60 11p13-q13.5	c	173	cSRL-3b3-tA cSRL-3b3-tZ	GCAGTATTTGGCTTTCTGTTCC AACCTAGGTGCCCATCAGTG	250 250	57
D11S1082	cSRL-3b8	0.85-0.93 11q23.3-q24.1	b	193	cSRL-3b8-tA cSRL-3b8-tZ	ACCCATTGTACCCATGTCAA CTCAAAAGAGCAGCCAAGGT	250 250	56
D11S1077	cSRL-3b10	0.60-0.85 11q14.1-q23.3	b	307	cSRL-3b10-tA cSRL-3b10-tZ	GATATGGACGTCTGCAAAACA TAAGGAAGGAAAGGGAGGGA	500 500	57
D11S1078	cSRL-3b11	0.60-0.85 11q14.1-q23.3	b	285	cSRL-3b11-tA cSRL-3b11-tZ	TTTGGAGCTGGACAGTGTTG TTCTCTATTGTGTCCCCTCCC	833 833	56
D11S1079	cSRL-3b12	0.33-0.50 11p11.2-q13.1	b	161	cSRL-3b12-tA cSRL-3b12-tZ	TGCAATGACGAAGCCCTACTG AAGGCACTTCCCTGCTCC	250 250	56
D11S1003	cSRL-3c4	0.93-1.00 11q24.1-q25	b	151	cSRL-3c4-tA cSRL-3c4-tZ	GGGAAAGAAGCCCCATAAG AAATTCTCCTGATCGACCTCA	833 833	56
D11S1084	cSRL-3d1	0.33-0.50 11p11.2-q13.1	b	152	cSRL-3d1-tA cSRL-3d1-tZ	TCAGTTCCTTCTTGGCACG CATGCCACATTTGCCATTTA	250 250	56
D11S1087	cSRL-3d3	0.50-0.60 11q13.1-q14.1	b	146	cSRL-3d3-tA cSRL-3d3-tZ	GAACAATATAACCATGCAACCC TCTACACTACTCCTTTTTCCCG	833 833	48
D11S1088	cSRL-3d5	0.00-0.05 11p15.5-p15.4	b	151	cSRL-3d5-tA cSRL-3d5-tZ	TGAAAGACAATTAGGTGAACCC TTCCTGGGCCATCTCATC	833 833	56
D11S1089	cSRL-3d6	0.00-0.05 11p15.5-p15.4	b	267	cSRL-3d6-tA cSRL-3d6-tZ	CCCCCATGATGACGTTAGTG GTGCGTTTTTGGCAGGATAT	250 250	56
D11S1085	cSRL-3d11	0.33-0.50 11p11.2-q13.1	b	255	cSRL-3d11-tA cSRL-3d11-tZ	AAGGATGGTTCAAGATCCCC CTGCCCCCCAAGTTACTGTA	250 250	56

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S1086	cSRL-3d12	0.24-0.33 11p13-p11.2	b	276	cSRL-3d12-1A cSRL-3d12-1Z	GGCAAGCACACCAACATAAGC AATTTGGCTCTGGGATTCCT	250 250	56
D11S1090	cSRL-3e1	nd		264	cSRL-3e1-1A cSRL-3e1-1Z	AGGTGAACTACCATGGCCAG GGGGAGCATGTACTTGCAGA	250 250	56
D11S1093	cSRL-3e2	0.33-0.50 11p11.2-q13.1	b	147	cSRL-3e2-1A cSRL-3e2-1Z	TTTACTTTGTTGTGATAAGACAACCC TGAATAATGTGGACACCAGAGC	833 833	56
D11S1094	cSRL-3e4	nd		206	cSRL-3e4-1A cSRL-3e4-1Z	ACCATCACCACCATCCATCT CAATGGAATATTGCTTAGCCTT	833 833	56
D11S1095	cSRL-3e5	0.00-0.05 11p15.5-p15.4	b	179	cSRL-3e5-1A cSRL-3e5-1Z	CCTTAGAAGCCAAAGCTCCC CATTATCGTGAGGCTGAAAGC	833 833	56
D11S1096	cSRL-3e8	0.85-0.93 11q23.3-q24.1	b	253	cSRL-3e8-1A cSRL-3e8-1Z	TGCCCCAAGCTCTACACTG GGTGTCGAGATGCAGCTGTA	250 250	56
D11S1091	cSRL-3e10	0.85-0.93 11q23.3-q24.1	b	250	cSRL-3e10-1A cSRL-3e10-1Z	CCACAGAGTGCATCATCCAT TCAGTGTTGTAAGTGGTGGGG	500 500	56
D11S1092	cSRL-3e12	0.33-0.50 11p11.2-q13.1	b	245	cSRL-3e12-1A cSRL-3e12-1Z	AGGCAAGTCTCCCATTTTTG TGCGCTAAATGGAAACACTG	250 250	56
D11S1099	cSRL-3h2	0.05-0.24 11p15.4-p13	b	257	cSRL-3h2-1A cSRL-3h2-1Z	TCTTCATGCAAGCCAAACAG CAGCCAGCAATTGTGTGG	833 833	56
D11S1100	cSRL-3h3	0.24-0.33 11p13-p11.2	b	250	cSRL-3h3-1A cSRL-3h3-1Z	TTGGTGCAAAAGTAATTGCG GAAGAAATGCACACAAATCCC	833 833	56

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S1101	cSRL-3h4	0.50-0.85 11q13.1-q23.3	b	250	cSRL-3h4-tA	AAACAGAAAACCAAACACTGCA	833	56
					cSRL-3h4-tZ	TGACATTAATGTGCAGCATGTG	833	
D11S1102	cSRL-3h6	0.24-0.33 11p13-p11.2	b	159	cSRL-3h6-tA	GACCATGCTTCGAGTCAACA	833	56
					cSRL-3h6-tZ	TAACGCCCTCATCATCTTGG	833	
D11S1103	cSRL-3h7	0.24-0.33 11p13-p11.2	b	252	cSRL-3h7-tA	CCTGCATCCATCCTTCAACT	250	56
					cSRL-3h7-tZ	ACGTGCAGAAGCTGGTGAGT	250	
D11S1104	cSRL-3h8	0.60-0.85 11q14.1-q23.3	b	255	cSRL-3h8-tA	AATTCACACATTCGAATTGAAA	833	56
					cSRL-3h8-tZ	AGCAGATCCACAATGCAGGT	833	
D11S1097	cSRL-3h10	nd		188	cSRL-3h10-tA	CCTCTGCCCTTTCCCATGTTA	833	56
					cSRL-3h10-tZ	CCTATGTTAAACCCCTGCCA	833	
D11S1098	cSRL-3h11	0.05-0.24 11p15.4-p13	b	245	cSRL-3h11-tA	AACAGCAAAAAACTAAAGCAATCC	833	56
					cSRL-3h11-tZ	GCGCTTTTAACTACAGGTACATTT	833	
D11S1105	cSRL-4a4	0.50-0.60 11q13.1-q14.1	b	157	cSRL-4a4-tA	TTGGATTGAAAGAACTTCGTCT	833	56
					cSRL-4a4-tZ	AAAGTCACAAGGACAATTTTTGC	833	
D11S1106	cSRL-4a6	0.05-0.24 11p15.4-p13	b	165	cSRL-4a6-tA	TTATCAGCAGCATGAAAACAGA	833	56
					cSRL-4a6-tZ	CATGCACCTTCAGCATGATC	833	
D11S1107	cSRL-4a7	0.24-0.33 11p13-p11.2	b	190	cSRL-4a7-tA	GAGATGAGAAAATCAAGGCTCG	250	56
					cSRL-4a7-tZ	GCACTCCGGTTGTTTCCATA	250	
D11S1108	cSRL-4a9	0.05-0.33 11q23.3-q24.1	b	150	cSRL-4a9-tA	CCACCACCACACTCCCAGT	250	56
					cSRL-4a9-tZ	GTCAGTTGCAGACACTTAGCC	250	
D11S1109	cSRL-4b1	0.50-0.60 11q13.1-q14.1	b	255	cSRL-4b1-tA	TTATGGGCTGTACTAAAGCGTT	250	56
					cSRL-4b1-tZ	TAGACACCATGCTTTATGCTGC	250	

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FL/ar range bands	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S1112	cSRL-4b3	0.05-0.24 11p15.4-p13	b	150	cSRL-4b3-1A cSRL-4b3-1Z	CCCTATCATTGGCAGCTTGT ATAAAACCAGCATCTTTTTCCC	500 500	56
D11S1113	cSRL-4b4	nd		249	cSRL-4b4-1A cSRL-4b4-1Z	CCTTTCCAGCTCACCTTCTG ACCCTGGCAGTGTCTAGGG	250 250	56
D11S1114	cSRL-4b5	0.05-0.24 11p15.4-p13	b	123	cSRL-4b5-1A cSRL-4b5-1Z	CCTTTTCACGTCCCTGACTC AAATGAGAAATACCCAGACTGG	833 833	56
D11S1115	cSRL-4b6	0.24-0.33 11p13-p11.2	b	260	cSRL-4b6-1A cSRL-4b6-1Z	AGCAAGACTCTGTCTCGAAAA AAAGTGCAAATAAGTAAGCCCA	833 833	56
D11S1116	cSRL-4b7	0.93-1.00 11q24.1-q25	b	203	cSRL-4b7-1A cSRL-4b7-1Z	ATGTAGCAACTGGGGCAGTC TGAAAGCAGTGAGTTTACGGG	833 833	56
D11S1117	cSRL-4b8	0.60-0.85 11q14.1-q23.3	b	251	cSRL-4b8-1A cSRL-4b8-1Z	TCTCTCATTTCAGCAGCATCA TCGTTCTTCTCCTTTGCCC	833 833	52
D11S1110	cSRL-4b10	0.50-0.85 11q13.1-q23.3	b	205	cSRL-4b10-1A cSRL-4b10-1Z	TCCCCCAGACTGTCTCTTG AGATCTAGGCTGGGACCAG	250 250	56
D11S1111	cSRL-4b12	0.85-0.03 11q23.3-q24.1	b	317	cSRL-4b12-1A cSRL-4b12-1Z	TTGAAGACCCCCAGAGAAGA CTAGAAGCAAGAGGGACTCAGC	833 833	58
D11S1120	cSRL-4c2	0.05-0.24 11p15.4-p13	b	251	cSRL-4c2-1A cSRL-4c2-1Z	CAATGCTCTCTCTCTATGGATTC TAACATGGTGGAAGGCATCA	833 833	56
D11S1121	cSRL-4c3	0.24-0.33 11p13-p11.2	b	213	cSRL-4c3-1A cSRL-4c3-1Z	TCGTTTTTGTACCAGAACAGAA TGGTGGCAGGGCACCTGTA	500 500	56

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpier range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S1122	cSRL-4c4	0.05-0.24 11p15.4-p13	b	193	cSRL-4c4-tA cSRL-4c4-tZ	ATGAAACTAAAAGTTTGTTTCAAAAGA TCCTCTAGATTTTCTCTTCTACTGGC	500 500	56
D11S1123	cSRL-4c6	0.24-0.33 11p13-p11.2	b	211	cSRL-4c6-tA cSRL-4c6-tZ	GCAAATCCCTGATTTACAGAGG AATTTTTTGGGGCAAACTGG	500 500	56
D11S1124	cSRL-4c8	0.93-1.00 11q24.1-q25	b	179	cSRL-4c8-tA cSRL-4c8-tZ	TCTATAACCCTGCCCACAGG AATGTGCTGGCAAGACAGCT	250 250	56
D11S1118	cSRL-4c10	nd		156	cSRL-4c10-tA cSRL-4c10-tZ	ACATAATGTGTGCCCATGGA ACGTAGTGATGAAGCCTGGC	250 250	
D11S1119	cSRL-4c11	0.85-0.93 11q23.3-q24.1	b	206	cSRL-4c11-tA cSRL-4c11-tZ	GGTTCATTTCAGTGCCCTGT CAGACAAGGACGGCCTCTAC	250 250	
D11S1125	cSRL-4d1	0.24-0.33 11p13-p11.2	b	179	cSRL-4d1-tA cSRL-4d1-tZ	CCTCTTAGGCTTATTGTGGGA TTCACTCTGTTGTGCTACCAAA	250 250	56
D11S1127	cSRL-4d4	0.33-0.50 11p11.2-q13.1	b	201	cSRL-4d4-tA cSRL-4d4-tZ	ATGAATTAGCCATTGGGCTG CCAGTGGAGGCCAGAGTAAA	500 500	58
D11S1128	cSRL-4d6	0.05-0.24 11p15.4-p13	b	139	cSRL-4d6-tA cSRL-4d6-tZ	TGCAAAATTTACTTGATTCCCAA TAACACATTATTTCTCTTTCACCATATG	500 500	56
D11S1129	cSRL-4d7	nd		171	cSRL-4d7-tA cSRL-4d7-tZ	CAAGACCCTGTCTCAAAAAAGC AAAAAAAAAATGTTGGTGGGG	250 250	56
D11S1126	cSRL-4d10	0.50-0.85 11q13.1-q23.3	b	257	cSRL-4d10-tA cSRL-4d10-tZ	TCCCCTTCAGCTATGCTAAA AGCTGCTGGCAAAAACAAAT	250 250	56
D11S1130	cSRL-4e1	0.05-0.24 11p15.4-p13	b	173	cSRL-4e1-tA cSRL-4e1-tZ	TCCTTTCGAAATTTTTTCCG AGTAGACATCGGATTGGGCC	500 500	56

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S1131	cSRL-403	nd		210	cSRL-403-1A cSRL-403-1Z	ACCAAAAGCAGTTTTGAGAAGG TTTCAAAAAACCAAGCTTTTCA	833 833	56
D11S1132	cSRL-404	nd		169	cSRL-404-1A cSRL-404-1Z	CTGTGCTTTCATGTTCATTGC ATTCACGTTGTTGTAGGTGACA	500 500	56
D11S1133	cSRL-405	0.60-0.85 11q14.1-q23.3	b	170	cSRL-405-1A cSRL-405-1Z	TCCTCCCCAAAACCCTTAAC ATTATGTCTCTGTCAGGTCCCC	833 833	56
D11S1134	cSRL-406	nd		163	cSRL-406-1A cSRL-406-1Z	CTCCCAAGATTTCAGAGGATG GCTCCACTAGGCAGTGCC	250 250	56
D11S1135	cSRL-408	0.60-0.85 11q14.1-q23.3	b	341	cSRL-408-1A cSRL-408-1Z	TGAGACTTGCCCCTCTACATCA TTCAGGCATTCACATTCTGC	250 250	54
D11S1136	cSRL-409	0.60-0.85 11q14.1-q23.3	b	259	cSRL-409-1A cSRL-409-1Z	GGATTCCCAGGTGAATGC ATAATTTCCATGCCCCCTTC	500 500	56
D11S1138	cSRL-412	0.33-0.50 11p11.2-q13.1	b	161	cSRL-412-1A cSRL-412-1Z	CACCCCATTGTCAAAAACC ACAGCTAAAGCAACGTTTTGC	250 250	54
D11S1139	cSRL-413	0.33-0.50 11p11.2-q13.1	b	161	cSRL-413-1A cSRL-413-1Z	ACTAGTATCGGCCAGCCAGA TCGCACCTACAAAGAGCCTT	833 833	54
D11S1140	cSRL-416	0.33-0.50 11p11.2-q13.1	b	260	cSRL-416-1A cSRL-416-1Z	GAGTCCAGATTTGAAGAGCAGG TGATTTGAGGAATAGAATGGCA	833 833	54
D11S1141	cSRL-418	0.05-0.24 11p15.4-p13	b	195	cSRL-418-1A cSRL-418-1Z	GCACTTCGGAGCCATCAC GGAGGGAAATAGTGGGGGTA	500 500	54

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S1142	cSRL-4i9	0.50-0.60 11q13.1-q14.1	b	196	cSRL-4i9-1A cSRL-4i9-1Z	GAGGTGGTGGGGAACCTGTA AGAGGGAGGAAGAACACCTT	250 250	54
D11S1137	cSRL-4i12	0.60-0.85 11q14.1-q23.3	b	323	cSRL-4i12-1A cSRL-4i12-1Z	AAATCTAGGCATCATGGCTCA GCTTCCATGCCAGGAAGAT	250 250	54
D11S1143	cSRL-4g1	0.24-0.33 11p13-p11.2	b	255	cSRL-4g1-1A cSRL-4g1-1Z	TGGTAGGTACTGCTCCTGGG CAGAATTGCCCAGCCTAATC	500 500	54
D11S1146	cSRL-4g3	nd		286	cSRL-4g3-1A cSRL-4g3-1Z	TGGGCTCTTTATTCTGTTCCA ATACTGCCCAAAGCAGTGTACA	500 500	56
D11S1147	cSRL-4g4	0.85-0.93 11q23.3-q24.1	b	263	cSRL-4g4-1A cSRL-4g4-1Z	GGAATCACTGAGTTAGAAGGAATAGG TTTGTAGCACTCTATCCCTCAAAACC	500 500	56
D11S1148	cSRL-4g8	0.24-0.33 11p13-p11.2	b	219	cSRL-4g8-1A cSRL-4g8-1Z	CCAAACCCAATTAGAAATCAG CACTGTTCCCAGCCTGTTTC	500 500	56
D11S1149	cSRL-4g9	0.05-0.24 11p15.4-p13	b	273	cSRL-4g9-1A cSRL-4g9-1Z	ACTAAATGAGCAAAGCCCCA AGAATCAAACCTATTCCCTCTG	500 500	56
D11S1144	cSRL-4g10	nd		154	cSRL-4g10-1A cSRL-4g10-1Z	AACTTGTGTCACAGGGGTTG AATGCATGCACACAGAGG	500 500	56
D11S1145	cSRL-4g11	0.05-0.24 11p15.4-p13	b	251	cSRL-4g11-1A cSRL-4g11-1Z	ACAAACCTTCACATAGAAGCCC TTTCTGAAAACAGTTGCCACC	250 250	54
D11S1152	cSRL-4h5	0.05-0.24 11p15.4-p13	b	154	cSRL-4h5-1A cSRL-4h5-1Z	ATGAAGGCCCTAGTCCCCC TGCCCTGTCCCTTTAGAG	833 833	56
D11S1153	cSRL-4h8	0.93-1.00 11q24.1-q25	b	262	cSRL-4h8-1A cSRL-4h8-1Z	AAGAGGACGACCATGTGAGC TTTTTGTGACTGCCCCCTTTC	250 250	56

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S1150	cSRL-4h11	0.24-0.33 11p13-p11.2	b	254	cSRL-4h11-1A cSRL-4h11-1Z	CCTTGTATTCAGTGCCTGAGC AGTGGACACTTCCCAGGATG	250 250	56
D11S1151	cSRL-4h12	0.85-0.03 11q23.3-q24.1	b	154	cSRL-4h12-1A cSRL-4h12-1Z	ACTCATTCATGGAACCCCAA CATCATGCTAGGCACTGACTG	250 250	56
D11S1154	cSRL-5a1	0.05-0.24 11p15.4-p13	b	151	cSRL-5a1-1A cSRL-5a1-1Z	GTTCCTGCAACTTCATTTTCC CTGGCAACCTGATAAATCTG	833 833	56
D11S1157	cSRL-5a2	0.50-0.85 11p13.1-q23.3	b	305	cSRL-5a2-1A cSRL-5a2-1Z	TAAAACAAGAGCAAACCAAACG GTTCAATCTTGGGAGCCTGT	250 250	56
D11S1158	cSRL-5a3	0.50-0.85 11q13.1-q23.3	b	277	cSRL-5a3-1A cSRL-5a3-1Z	ATTTTAACCTCACTGTCCCTCA TTCAGACAGTGCTGCCCAC	250 250	56
D11S1159	cSRL-5a4	0.60-0.85 11q14.1-q23.3	b	250	cSRL-5a4-1A cSRL-5a4-1Z	ACAGTGTTGAATCCGGATTTC CTGTGCTATGTTCTAAGGACC	250 250	56
D11S1160	cSRL-5a5	0.60-0.85 11q14.1-q23.3	b	164	cSRL-5a5-1A cSRL-5a5-1Z	TCATTACCATTTTTATGTGCGG TCGATAAATCGAAATTCCTCA	500 500	56
D11S1161	cSRL-5a6	0.93-1.00 11q24.1-q25	b	155	cSRL-5a6-1A cSRL-5a6-1Z	GACAGTGGAGCCAATGTATTTT CAAAGACTTTATCTTGGCCC	833 833	52
D11S1162	cSRL-5a7	0.24-0.50 11p13-q13.1	b	204	cSRL-5a7-1A cSRL-5a7-1Z	AAGTCAGCGACCTTTTTGAGC ATAAATCCCCTTGATTAGGTGC	250 250	56
D11S1163	cSRL-5a9	0.85-0.03 11q23.3-q24.1	b	258	cSRL-5a9-1A cSRL-5a9-1Z	CCTGAGCCTGACTTAGCCTC TCCAGATGGGAAAACTGAGC	250 250	56

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	Filter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S1155	cSRL-5a10	0.24-0.33 11p13-p11.2	b	285	cSRL-5a10-1A cSRL-5a10-1Z	GAAAAGATAAACCCAACGTCA ACGCCACACATTTTTACCC	833 833	56
D11S1156	cSRL-5a11	0.50-0.60 11q13.1-q14.1	b	290	cSRL-5a11-1A cSRL-5a11-1Z	GTCACTCACTCATTTGAGGTCC TCGAGTGGACTGAACTGAAGG	250 250	56
D11S1164	cSRL-5b1	0.93-1.00 11q24.1-q25	b	198	cSRL-5b1-1A cSRL-5b1-1Z	CCATGATTCCAAAACAACCC CCCTCTCTCCTTACCTCCAA	500 500	56
D11S1165	cSRL-5b7	0.05-0.24 11p15.4-p13	b	151	cSRL-5b7-1A cSRL-5b7-1Z	AAATATCACTCCCAGGCCAA AAACTGCAATTACTTTTCCACC	500 500	56
D11S1166	cSRL-5b8	0.05-0.24 11p15.4-p13	b	150	cSRL-5b8-1A cSRL-5b8-1Z	TGCTTGCCAAATCTATATCG AAATATCCCCAGGATTTCTTT	250 250	56
D11S1169	cSRL-5c2	0.50-0.85 11q13.1-q23.3	b	250	cSRL-5c2-1A cSRL-5c2-1Z	AAGGACCAGAGTGTGGGCT AGTCAGATGAAGCAAACCA	833 833	56
D11S1170	cSRL-5c3	0.05-0.24 11p15.4-p13	b	278	cSRL-5c3-1A cSRL-5c3-1Z	CTACATGACCTGCCTCCCTT TGATTGCCAATGGCAGATA	500 500	58
D11S1171	cSRL-5c5	0.24-0.33 11p13-p11.2	b	307	cSRL-5c5-1A cSRL-5c5-1Z	CAGAGCCACTACACTCTCATTG CAAGTGCAACAGTTGAGAACC	500 500	58
D11S1172	cSRL-5c8	0.00-0.05 11p15.5-p15.4	b	155	cSRL-5c8-1A cSRL-5c8-1Z	CCCCTGACTTAAATAAATCCTG AGTTTTGGTAACATGCTGTGG	500 500	56
D11S1173	cSRL-5c9	0.60-0.85 11q14.1-q23.3	b	242	cSRL-5c9-1A cSRL-5c9-1Z	TTGTTCAACTATTACTGACACTACTGC CAGGCTGATGCATACTTACTTGT	500 500	56
D11S1167	cSRL-5c10	0.60-0.85 11q14.1-q23.3	b	261	cSRL-5c10-1A cSRL-5c10-1Z	ACATGAGCCACCATGCCT CCTTGCCAAATCTGCTCAAC	250 250	56

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S1168	cSRL-5c11	0.05-0.24 11p15.4-p13	b	269	cSRL-5c11-tA cSRL-5c11-tZ	GCGGCATGCATTCTGACTAA CAACAGAACCGCCTTTTGCAG	833 833	56
D11S1176	cSRL-5d2	0.33-0.50 11p11.2-q13.1	b	252	cSRL-5d2-tA cSRL-5d2-tZ	GTCCTGCCCTTCATTACTGC CATGCTGGACACTACTCTGACC	500 500	58
D11S1177	cSRL-5d3	nd		160	cSRL-5d3-tA cSRL-5d3-tZ	AGAAACTCACGTATGCCAAGG TTTAGTGCTTTGACACATTTGC	250 250	56
D11S1178	cSRL-5d5	0.60-0.85 11q14.1-q23.3	b	312	cSRL-5d5-tA cSRL-5d5-tZ	GTGTTCTTAAAGGGCTCTGCC CATGCATTCTTCCTGTCGG	500 500	56
D11S1179	cSRL-5d6	0.33-0.50 11p11.2-q13.1	b	287	cSRL-5d6-tA cSRL-5d6-tZ	ATACCCCATGGTTGCCTAAG TGACAATATCAAGTGTGGCCA	500 500	56
D11S1180	cSRL-5d7	0.60-0.85 11q14.1-q23.3	b	200	cSRL-5d7-tA cSRL-5d7-tZ	GGAAGCAGTATTACACTGAGAACT TAATGTGTTTTAGCTGCAAGGCA	500 500	56
D11S1181	cSRL-5d8	0.05-0.24 11p15.4-p13	b	211	cSRL-5d8-tA cSRL-5d8-tZ	CAGCAGAAATCTCTGTGACACC TATTGTGAGACACGTCTGGGG	250 250	56
D11S1182	cSRL-5d9	0.50-0.60 11q13.1-q14.1	b	297	cSRL-5d9-tA cSRL-5d9-tZ	TTTAGGAATCAAATCTGATGCC AGGGTGGATTGCTGAACTTG	500 500	58
D11S1174	cSRL-5d10	0.50-0.85 11q13.1-q23.3	b	251	cSRL-5d10-tA cSRL-5d10-tZ	TCCTTCTCAAGTCTCTCCTTCC GCAAGCCATATCTAGCATTTTT	250 250	56
D11S1175	cSRL-5d11	0.93-1.00 11q24.1-q25	b	186	cSRL-5d11-tA cSRL-5d11-tZ	AGCAGTGGCCTAAACCATTG CACACTGTGGACTGGATTGG	500 500	58

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range based(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S1183	cSRL-5e1	0.33-0.50 11p11.2-q13.1	b	287	cSRL-5e1-1A cSRL-5e1-1Z	AGACTGCCTTTCCCTGCC AGCAAATTGGAGGCTGGTTT	250 250	58
D11S1185	cSRL-5e2	0.93-1.00 11q24.1-q25	b	334	cSRL-5e2-1A cSRL-5e2-1Z	AAAAGGTCCTGGCATACAGC AACTAATTGGTCAGGCCACG	833 833	58
D11S1186	cSRL-5e3	0.05-0.24 11p15.4-p13	b	287	cSRL-5e3-1A cSRL-5e3-1Z	CCCAATTGTCTCCCTAAATCC AATCTGCATGGTCAAAGATGG	833 833	58
D11S1187	cSRL-5e4	0.05-0.24 11p15.4-p13	b	256	cSRL-5e4-1A cSRL-5e4-1Z	CATTGTTGCCACAGAGATGG CACTTCCACATTTGCAAATCA	833 833	58
D11S1188	cSRL-5e5	0.33-0.50 11p11.2-q13.1	b	269	cSRL-5e5-1A cSRL-5e5-1Z	CACAGAGCAGCAAACTCAAGG TCTGACTCCTTCCTGTCTAGGG	833 833	58
D11S1189	cSRL-5e6	0.50-1.00 11q13.1-q25	b	263	cSRL-5e6-1A cSRL-5e6-1Z	CTTCCCATTGAAACCCTTACG GCTCTATTCCTATCCGTTTCCC	833 833	52
D11S1184	cSRL-5e10	0.05-0.24 11p15.4-p13	b	211	cSRL-5e10-1A cSRL-5e10-1Z	TCAGGTCTCATGATATTGGGC TTAGCTGGCTGTCTGTCTTACC	500 500	56
D11S1190	cSRL-5f1	nd		254	cSRL-5f1-1A cSRL-5f1-1Z	CTGAGGGCATTTCCCTCTAA CCAGGCCCTTCATCCATC	250 250	56
D11S1193	cSRL-5f2	0.00-0.05 11p15.5-p15.4	b	209	cSRL-5f2-1A cSRL-5f2-1Z	CCTGGGTTAAAACACTCAATTC AGGAAGCTGATTTTCTCCTGC	833 833	56
D11S1194	cSRL-5f3	0.33-0.50 11p11.2-q13.1	b	215	cSRL-5f3-1A cSRL-5f3-1Z	CTCACAACCCTCCATGGC ACTAAACCACCACGGCAGG	250 250	56
D11S1195	cSRL-5f4	0.05-0.24 11p15.4-p13	b	230	cSRL-5f4-1A cSRL-5f4-1Z	AAGTAAAGCCCGCAAAAGCAA GCTTAAACTATTTTGGGGGACT	250 250	56

SUBSTITUTE SHEET (RULE 26)

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (11 different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S1196	cSRL-515	nd		273	cSRL-515-1A	ATACCTCAGGCACCTGCATC	250	58
					cSRL-515-1Z	CACAACTGGTCACCAACACC	250	
D11S1197	cSRL-516	nd		315	cSRL-516-1A	CCACCATCAAATGCCACCC	250	56
					cSRL-516-1Z	TGACGGGTCACCAGCAAT	250	
D11S1198	cSRL-517	0.60-0.85	b	270	cSRL-517-1A	CACTGGAGCAAAAGCCAAC	500	56
		11q14.1-q23.3			cSRL-517-1Z	AAGCCCATACCCTCTTCTCA	500	
D11S1199	cSRI-519	0.93-1.00	b	151	cSRL-519-1A	GCTTGCGCTGCCAGAGAC	500	56
		11q24.1-q25			cSRL-519-1Z	TTCAGGTCATCAAAATCTGACA	500	
D11S1191	cSRL-5111	0.60-0.85	b	201	cSRL-5111-1A	AAATGACAGTGGAGCCCACT	500	56
		11q14.1-q23.3			cSRL-5111-1Z	CACAGCCCTTGCAAATCAAA	500	
D11S1192	cSRL-5112	0.60-0.85	b	165	cSRL-5112-1A	GTTCCTTTTAGCTTCAGTCCTT	500	56
		11q14.1-q23.3			cSRL-5112-1Z	ACCAATCAATCCAATGCAAACC	500	
D11S1200	cSRL-5g1	nd		166	cSRL-5g1-1A	CCCCCCTCCTTTCCTACT	500	56
					cSRL-5g1-1Z	AGTCCAGAAGAGGAGATGTTC	500	
D11S1203	cSRL-5g2	0.05-0.24	b	293	cSRL-5g2-1A	TTTTTCAGACCATTTTATTTCAATC	833	56
		11p15.4-p13			cSRL-5g2-1Z	CAAATTTATACTACCCATGCCC	833	
D11S1204	cSRL-5g6	0.33-0.50	b	256	cSRL-5g6-1A	CCCCCCTCTTCATTACCTTT	833	56
		11p11.2-q13.1			cSRL-5g6-1Z	AGCTGACTCGACTCACTTTCC	833	
D11S1205	cSRL-5g8	0.05-0.24	b	258	cSRL-5g8-1A	CTCACCTCCCCTCTCTCAC	500	56
		11p15.4-p13			cSRL-5g8-1Z	CTACCTCGAGGGAGGGACTC	500	

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (11 different)	Filter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S1206	cSRL-5g9	0.05-0.24 11p15.4-p13	b	335	cSRL-5g9-1A cSRL-5g9-1Z	GGGCTTAAATACCACACAATGG TGAGGGGAGTGTTAAGGCAC	833 833	52
D11S1201	cSRL-5g10	0.24-0.33 11p13-p11.2	b	168	cSRL-5g10-1A cSRL-5g10-1Z	ACCTGCACATCCTGCACAT GCCATCTATTCCTGGAGAGAG	500 500	56
D11S1202	cSRL-5g11	0.33-0.50 11p11.2-q13.1	b	150	cSRL-5g11-1A cSRL-5g11-1Z	ACAAAAATGGCAACTAGACCA AACAACTCTGGAGGGCTGG	500 500	56
D11S1209	cSRL-5h2	0.33-0.50 11p11.2-q13.1	b	260	cSRL-5h2-1A cSRL-5h2-1Z	TCGATGACCAGCCAACAATA CCATGGGTCCAAGCAGAC	500 500	56
D11S1210	cSRL-5h3	0.85-0.93 11q23.3-q24.1	b	302	cSRL-5h3-1A cSRL-5h3-1Z	AAACCTGCCCAGCCCTAC TGTGTTATGGGCCATGGTC	250 250	56
D11S1211	cSRL-5h4	0.60-0.85 11q14.1-q23.3	b	273	cSRL-5h4-1A cSRL-5h4-1Z	CCAAAAACCAGTACATCATCATCCA CCACCAAATCCATGTTTTCA	833 833	56
D11S1212	cSRL-5h6	0.60-0.85 11q14.1-q23.3	b	207	cSRL-5h6-1A cSRL-5h6-1Z	TTTAGACCCACAAAAGCAGTT AAACTGCAAACTTGCAAAGTTC	833 833	52
D11S1213	cSRL-5h7	0.05-0.24 11p15.4-p13	b	271	cSRL-5h7-1A cSRL-5h7-1Z	GCTGTAAGAGCTCCCTGCTG GAGAGTCTGGACCAATCCCA	833 833	56
D11S1214	cSRL-5h9	0.24-0.33 11p13-p11.2	b	205	cSRL-5h9-1A cSRL-5h9-1Z	TGAAGCAAAACAGCGAATCC TTTTTGGGCATGTCTGATCA	833 833	56
D11S1207	cSRL-5h10	0.60-0.85 11q14.1-q23.3	b	320	cSRL-5h10-1A cSRL-5h10-1Z	ATCCAAGGTCCCTGATAAACTT GGAAAGCAAAAAGCGAAAACT	500 500	52
D11S1208	cSRL-5h12	0.05-0.24 11p15.4-p13	b	268	cSRL-5h12-1A cSRL-5h12-1Z	ACTCCTAGCCACCCCAATCT TACTGCCGTAGCTGCATGAC	500 500	56

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FISH range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S1216	cSRL-6a2	0.33-0.50 11p11.2-q13.1	b	182	cSRL-6a2-1A cSRL-6a2-1Z	CAGTCATTCCAACCTTCCAC CCAAGAGACAAAGCGTTCCT	833 833	56
D11S1217	cSRL-6a3	0.33-0.50 11p11.2-q13.1	b	267	cSRL-6a3-1A cSRL-6a3-1Z	CCCCTCCCAGTAGTTCAAGT CATGCCGTCCAAATCTCG	500 500	56
D11S1218	cSRL-6a4	0.85-0.93 11q23.3-q24.1	b	318	cSRL-6a4-1A cSRL-6a4-1Z	CCAGCACTCTTTCAGTCTCAA TCAGCATGCATTTATCCAGC	833 833	56
D11S1219	cSRL-6a5	0.33-0.50 11p11.2-q13.1	b	205	cSRL-6a5-1A cSRL-6a5-1Z	TAGCCTCAAATCAACCAAGTCA AGATGGGGTCTCACTGTTGC	833 833	56
D11S1220	cSRL-6a6	0.05-0.24 11p15.4-p13	b	180	cSRL-6a6-1A cSRL-6a6-1Z	CCTCTCCCAACAATCTCGTT AGTAGAAGCATAGCATTTGCCA	500 500	56
D11S1221	cSRL-6a7	0.33-0.50 11p11.2-q13.1	b	306	cSRL-6a7-1A cSRL-6a7-1Z	GTGTTTTTGAGACAGTCTCGCT CCTCAGAAGTTTGCTCGTTAAGA	500 500	56
D11S1215	cSRL-6a12	0.60-0.85 11q14.1-q23.3	b	220	cSRL-6a12-1A cSRL-6a12-1Z	ACTGCTTCTGACAGCAATTCA TTCCCAGGTCAAATATTGGTG	833 833	56
D11S1222	cSRL-6b1	0.05-0.24 11p15.4-p13	b	151	cSRL-6b1-1A cSRL-6b1-1Z	GCATCTGTGCAGTCTTCTGC TAGTGCGCAGTCAGTGCTTG	833 833	52
D11S1225	cSRL-6b2	0.33-0.50 11p11.2-q13.1	b	276	cSRL-6b2-1A cSRL-6b2-1Z	TGCCTGCTCTGCTTTTTAGC AAGTGTGCACTACCTTGTCACC	833 833	52
D11S1226	cSRL-6b7	0.50-0.50 11c13.1	b	306	cSRL-6b7-1A cSRL-6b7-1Z	CTGCATGCCCAGAGAAAAC ATACCCTCAGTGCCCAAGAG	250 250	56

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (11 different)	FISH range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	(primer) (nM)	Annealing temp.(°C)
D11S1227	cSRL-6b9	0.33-0.50 11p11.2-q13.1	b	205	cSRL-6b9-1A cSRL-6b9-1Z	CTCAAGCTCTCGTGTGACTTTGG GGAATGCTACTGGAGGATTCA	833 833	52
D11S1223	cSRL-6b10	0.33-0.50 11p11.2-q13.1	b	151	cSRL-6b10-1A cSRL-6b10-1Z	TAACAGCACACTCTATGCCACC AATCTGTAAAGTCAGCATCCCC	833 833	52
D11S1224	cSRL-6b12	0.60-0.85 11q14.1-q23.3	b	274	cSRL-6b12-1A cSRL-6b12-1Z	CAAGAACATGCCAATTAAGCC CACCACGTAACCTCAGCAGA	833 833	52
D11S1228	cSRL-6c1	nd		271	cSRL-6c1-1A cSRL-6c1-1Z	GCCAACCAAGAGGGAAAGTT AGAACTGCCTCTGTCCCTCTGACA	833 833	52
D11S1231	cSRL-6c2	0.50-0.85 11q13.1-q23.3	b	336	cSRL-6c2-1A cSRL-6c2-1Z	AGACATTGCCCTGCTTCCTG CCCAGTGCCAGCCTTGTC	250 250	56
D11S1232	cSRL-6c4	0.85-0.85 11q23.3	b	335	cSRL-6c4-1A cSRL-6c4-1Z	CCCATTCAGTTCAACCTTCTC CCTGCACATAACCCCATTGT	500 500	58
D11S1233	cSRL-6c8	0.50-0.60 11q13.1-q14.1	b	180	cSRL-6c8-1A cSRL-6c8-1Z	CCCTATTACTCAGCACCCACC CACTGCTTTATCCCCCCTAG	500 500	56
D11S1229	cSRL-6c11	nd		275	cSRL-6c11-1A cSRL-6c11-1Z	CAGCTCAGCCTAACCTGTGG GTATCAGTTAGCTTTTGCTCCC	500 500	56
D11S1230	cSRL-6c12	0.50-0.85 11q13.1-q23.3	b	256	cSRL-6c12-1A cSRL-6c12-1Z	CCACAGACTCCCCTTGGACAT GTTCCCCTGTAAGGAATGACC	500 500	56
D11S1234	ySL-A109g11	nd		180	SLA109g11-A SLA109g11-Z	GTGCATATTGAACCAGATGGCCCCC TGGACCTGCTTGGTACAACGACCGA	500 500	51
D11S1235	ySL-A109G4	0.85-0.85 11q23.3	b	106	SLA109g4-A SLA109g4-Z	GCACCAGTCCTGCCCTAGAAGATTGTGCTC TAGCTGCATCTGAAACCTGCAGGGGCTC	250 250	54

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S1236	ySL-A13E4	0.91-0.93 11q24.1	a	192	SLA13E4-A SLA13E4-Z	CGCCGTGTTAAAGCTGCAACATGTCAC ATCATTGGTTATGCTCATTATCTGTTCCAA	250 250	56
D11S1237	ySL-A77g8	0.91-0.92 11q24.1	ab	337	SLA77G8-A SLA77G8-Z	GTGTAGTACAAAGGCACACTG CTAGGGAGACGTGCCATTCAA	250 250	54
D11S1238	ySL-B215B10	0.58-0.55 11q13.5-q13.3	ab	185	SLB215B10-A SLB215B10-Z	ACTCCATGAAGGTTCCATGGCTGCA TCTGTGCCCATAGAAGGGACCTAGG	833 833	54
D11S1239	ySL-B34g9	0.85-0.87 11q23.3	ab	160	SLB34G9-A SLB34G9-Z	CCCTCCACTAGGCAAATGCTCTCTCATTCC TGGCATTGTCAAGAGTGGACCATGAGCATC	250 250	56
ACP2		0.31-0.40 11p11.2-p11.11	bc	444	ACP2-B ACP2-Y	CCTCCTCATAGTGCTGCTCC CATAGAAGTGCAAGCTGGATCC	250 250	58
ACRV1		nd		202	ACRV1-A ACRV1-Z	CTTGCTTGTTTTGACTCAGGC AGAGAAAGAGTTGGAGCAGGG	500 500	58
ACTN3		0.33-0.50 11p11.2-q13.1	b	164	ACTN3-A ACTN3-Z	CTGGACTACGTGGCCTTCTC TTCAGAAGCACTTGGCTGG	833 833	58
AHNAK		0.00-0.40 11p15.5-p11.11	c	159	AHNAK-A AHNAK-Z	TGCCAGATGTTGACATTAAAGG GCAGCTTCACATCCACTTCA	250 250	58
AN2		0.23-0.27 11p13	bc	870	AN2-A AN2-Z	TGAATGGGCCGAGTTATGAT TATCAGGTTCACTTCCCGCA	250 250	60
APOA1		0.85-0.87 11q23.3	abc	258	PCR-AI.1 PCR-AI.2	CACCCCGCAGACCTGCAAGC TCTAAGCAGCCAGCTCTTGCA	833 833	56

SUBSTITUTE SHEET (RULE 26)

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
BCL1		0.52-0.55 11q13.3	c	205	BCL1(1) BCL1(2)	GCTAATCACAGTCTAACCGA TTGCACTGTCTTGGATGCA	833 833	52
BDNF		0.23-0.24 11p13	bc	579	BDNF-A BDNF-Z	CAGGCCACTGCTGTTCCT ACCTTGTCCTCGGATGTTTG	833 833	56
C1NH		0.42-0.50 11q12.1-q13.1	bc	418	C1NH-A C1NH-Z	ATCCTGGAGGTTTCCAGCTT TGTCCAACAAATCACCTGGA	833 833	56
CALCA		0.00-0.05 11p15.2-p15.4	bc	425	CALCA-A CALCA-Z	CTCAGGCACAGTGGAACTGA GTCCCAGAGACATTCTATTCCG	833 833	56
CALCB		0.09-0.13 11p15.2-p15.1	bc	381	CALCB-B CALCB-Y	AGAGGCACCACTACCCGACT AGCACCGCAGCTCAGTTC	833 833	56
CALCP		0.09-0.13 11p15.2-p15.1	bc	150	CALCP-A CALCP-Z	CTTCCCCTCCACAGGTCTC CTGGAGCTCTGTCTCCTGCT	250 250	58
CAT		0.24-0.27 11p13	bc	563	CAT-A CAT-Z	AAACAGAATGCGATTCACACC ATTAAGCCATGACCGTGCTC	833 833	56
CCND1		0.48-0.50 11q13.1	bc	219	PRAD1-A PRAD1-Z	CAGGCTGTGTCCCTCTTCTC CAGAAGCTATTCCAATCATCCC	250 250	56
CD20		0.48-0.52 11q13.1	c	158	CD20(1) CD20(2)	CAGCAGGCATCTATGCACCCATCTG TCTGGCATATCCCTGTGGAGCCTTT	833 833	44
CD3D		0.87-0.89 11q23.3	abc	800	CD3(1) CD3(2)	AAGGTATTTTGGCCCAGTCAATCAAAGGT AGTCATACACCTTAACCAAGTGGTTTCC	833 833	55
CD44		0.24-0.27 11p13	bc	150	CD44-A CD44-Z	GCAGAATGTGGACATGAAGA ATGCTAAAAAAGATTCGCAATC	500 500	56

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	(primer) (nM)	Annealing temp.(°C)
CD5		0.48-0.50 11q13.1	bc	950	CD5-A CD5-Z	CAGCTCAATCATCTCCTACGG TGCCCCTGTAGAACTCCAC	833 833	56
CHRM1		0.42-0.59 11q12,1-q13.5	c	543	CHRM1-B CHRM1-Y	CAGTCACAGGCAACCTGCT GTCCTCCGTTCTCTGTCTCC	250 250	60
CHRM4		0.27-0.37 11p12-p11.2	c	417	CHRM4-A CHRM4-Z	CTCACTCTCTCATCACCCTGC TAGCCATTGTGGCATCATGA	833 833	50
CK2A		0.24-0.50 11p13-q13.1	b	184	CK2A-A CK2A-Z	ACAACGTCTTACCCCCAGCT TTCCACCACATGTGACTCGT	500 500	58
CLG		0.68-0.70 11q21	bc	461	CLG-A CLG-Z	TTATGACCATCAGAACCAGCC CTGTGACACACCACACCCC	833 833	56
CLG1		0.68-0.78 11q21-q22.3	bc	171	CLG1-A CLG1-Z	TAACCCCAGGTACCCATGAA TCTGTCTTCACACGGCCTC	500 500	56
CNTF	CNTF1	0.41-0.52 11q11-q13.1	a	486	CNTFS CNTFALPHA	GTGAAGCATCAGGGCCTGAAC CATTTTCTTGTTGTTAGCAAT	250 250	56
COX8		0.48-0.59 11q13.1-q13.5	c	241	COX(1) COX(2)	GCCCTTACCTCCTGCTTCGTGACCTTCCTC GAGGTCATGATTGCAGAAGACGTGACTGCA	833 833	65
CRYAB	CRYA2	0.75-0.82 11q22.3-q23.1	bc	441	CRYA2-A CRYA2-Z	CGTTTGCCACGTTTCCACAC CTGGTACAGAGTCAGGAATCCC	833 833	56
D11S149		0.33-0.37 11p11.2	bc	135	D11S149(1/1) D11S149(2/2)	AGCCAGATAACTTACAAAATCCTA TGCTGGCAAACTCGTTTCTGTCCT	833 833	60

SUBSTITUTE SHEET (RULE 26)

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
D11S325		0.24-0.27 11p13	bc	750	D11S325.PCR1.1 D11S325.PCR1.2	GACAGACACACAGGAGAGAATGAATATAT CCAGTGCAGCAGAAGCAAAGCGCGG	250 250	56
D11S35		0.70-0.78 11q22.1-q22.3	bc	165	780 781	ACAATTGGATTACTACTACC TGTATTTGTATCGATTAACC	833 833	52
D11S367		0.85-0.84 11q23.3	abc	203	D11S367-A D11S367-Z	GACCTGCCTGTGTTTGTGCTCTGTGT TTCTTCACTGCAGGTAGAAGCAGCCC	250 250	56
D11S384		0.78-0.78 11q23.1	abc	273	cj193 5' cj193 3'	GACGGCTAACTGATGTCTAC ATTAGCCAAGCAGTTGCCAGC	500 500	56
D11S419		0.05-0.24 11p15.4-p13	bc	112	Mld58CA Mld58GT	CTCATTTGAAGACTGCAGCA AGGGCTTCCTGTCCATCTA	833 833	56
D11S420		0.89-0.92 11q23.3-q24.1	abc	200	506 507	AGTTACACCGGTTCTGCAGA GATTAATGATAGTGCTATCC	833 833	50
D11S490		0.85-0.87 11q23.3	abc	150	22314 41919	CACAAACATTGGCGCAT TTCTGGGTCACGGTGCTTCA	160 160	50
D11S528		0.85-0.91 11q23.3	bc	73	42026 42027	AATGGTGTCCCCACACATGT TCCTACCTACCGAGCTTAAA	833 833	55
D11S533		0.48-0.59 11q13.1-q13.5	c	500	D11S533.PCR1.1 D11S533.PCR1.2	GCCTAGTCCCTGGGTGTGGTC GGCGGTCTGCGAACATGTCCCC	500 500	58
D11S534		0.50-0.59 11q13.1-q13.5	b	200	91240 92005	ATATGCAAACTCTCCGTACT GCAACCATGGAGAGTCTGCA	500 500	56
D11S790		0.55-0.55 11q13.3	a	450	c11q-3h10-uA c11q-3h10-uZ	GTAGTTGTGCGCCAGGGTAA AAACAATCCCTCCTCGCC	500 500	56

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	(primer) (nM)	Annealing temp.(°C)
D11S860		0.00-0.13 11p15.5-p15.1	c	154	BS48L BS48R	TAGTATTGCCCATAGAAGAAGC GCAACACGTACACACTCAGACA	833 833	55
D11S861		0.05-0.24 11p15.4-p13	b	154	A136-1 A136-2	CTCAAACCAACTGAAAAGGACA AAGCTCCATTGTGTTCTGCC	250 250	55
D11S862		0.60-0.85 11q14.1-q23.3	b	152	MS7-1 MS7-2	TACCATATTAAATCACCCACATGG GGAATCAGCATTCAAACTCTGG	833 833	58
D11S863		nd		133	MS20-1 MS20-2	GCAACATGGTAAGAGTCCAGC GTAGCTGGGACCAGAGGGAT	833 833	60
D11S865		nd		170	E137-1 E137-2	CTTTTTGTTGCCCATTGCTT GTAAGGATAAAACTCTGCAAGGC	833 833	58
D11S869E		nd		277	AA12A AA12B	AGTGATCTCTGAGAAAAGGG ATATTGTGAATGACTAGGG	500 500	55
D11S870		0.50-0.85 11q13.1-q23.3	bc	154	349 350	ATTTTGGATGACCCAAGCCT ATCTGTATATATGTGTACCTG	500 500	55
D11S872		0.93-1.00 11q24.1-q25	bc	158	201 401	GATCTCTCTCTCAAACACAAAC AAGTCTTCAGCTTTATCAAC	833 833	56
D11S873		0.50-0.85 11q13.1-q23.3	bc	176	489 490	CCTCGTTTAGAATAATACCT ATAATGTACTGTGATAAATGCT	833 833	54
D11S074		nd		156	583 584	CGTTTAAAAGTCAGCCCTC AATCATTTTCAAGCATAGGC	500 500	56

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range bands	Map tech.	Product size (bp)	Primer name	Primer sequence	(primer) (nM)	Annealing temp.(°C)
D11S875		0.05-0.24 11p15.4-p13	bc	103	599 600	ACTGTCCTCTCATCCTACTG TACAGACCTGAGTTTGTAGC	833 833	56
D11S876		0.60-0.85 11q14.1-q23.3	bc	216	701 702	TGGAGATGTGCCATAGAGGT TCAGCAAAACTGCCCTGACG	250 250	56
D11S943E		0.85-0.93 11q23.3-q24.1	bc	103	307 308	CAACGGAACAGGAGTCCTT GCTTCTGACACTTCAATCTA	833 833	56
D11S944E		0.50-0.85 11q13.1-q23.3	bc	117	139 140	CTTTGAAGGAAGTGATTTCC TAGGCCCACCTCCAGTTCAT	833 833	56
D11S950E		0.93-1.00 11q24.1-q25	b	242	os100111-A os100111-Z	GAATGTACCCTTAAAGGATGG CCAGTTCAATTTACCCTTCAGG	250 250	57
D11S952E		0.33-0.50 11p11.2-q13.1	bc	115	339 340	TTTCCGAAGAAGCAGTTTG ATCAGCCTAGAGCCCTGACT	833 833	56
D11S953E		0.60-0.85 11q14.1-q23.3	b	136	os100016-A os100016-Z	GCTACTACCATCCCATGGGT TTACGATAATGACATTTCTTCTGG	250 250	56
D11S956		0.33-0.50 11p11.2-q13.1	bc	247	sMSH3A sMSH3B	GATCACTAATTAGCCAGACTCTAGG CGTTTTGGAGCTTAAGAGC	250 250	56
D11S964		nd		291	UT544a UT544b	ACTTCAGCCTCGGTGACAC TGTTCTGCCTCTGTTGTTAC	833 833	56
D11S968		0.93-1.00 11q24.1-q25	b	150	AFM109xc3a AFM109xc3m	CGCTCTTGTAGTTCTTATCTCCT AAGCCCATCCTCGAC	833 833	52
D11S969		0.93-1.00 11q24.1-q25	b	150	AFM205vf10a AFM205vf10m	TTGATTTGGAAGATTTTCAC CGGGCAGAATGGGTAT	833 833	49

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
DRD2		0.79-0.82	a c	283	JHE20	CTCCTTCTGCTGGGAGAGCTTCCTATGGCT	500	65
		11q23.1			JHE21	CGAGCCCAGGAGCTGGAGATGGAGATGGAG	500	
DRD4		0.00-0.03	c	478	DRD4-A	AGCTTATAGAGGCTGCTGGG	250	50
		11p15.5			DRD4-Z	TTTTTGGATGGATTTGTTTCC	250	
FDX1		0.60-0.59	bc	461	FDX1-A	CTGTTCAGTGCTCCTCAAACA	833	56
		11q14.1-q13.5			FDX1-Z	GTATGCCCCTAAACAAGCCA	833	
FGF4		0.52-0.55	c	310	HST(1)	CAGTGTTGGAACTGCGGGCTTGAGCCTGGA	833	52
		11q13.3			HST(2/2)	ACGACAATCTCCTTCAACGTGCACTCATCG	833	
FLI1		0.93-0.94	a c	377	EWIS	AGTCTCCTAGCATCTTGTGAGTTGCAT	250	58
		11q24.1-q24.3			EW2A	GTGGTAGCTGAGTGCATACTGTCCAC	250	
FOLR1		0.52-0.60	bc	300	FOLR1-A	CGGCATTTCATCCAGGAC	250	57
		11q13.3-q14.1			FOLR1-Z	CGTGTAGGAGGTGCCACAAT	250	
FSHB		0.23-0.26	a c	409	FSHB-A	ATAACGACCCAGCCAGGC	833	58
		11p13			FSHB-Z	CATCATGGAGTCCTGCCC	833	
FTH1		0.48-0.59	c	284	FTH1-A	TGCTTCAACAGTGCTTGGAC	250	56
		11q13.1-q13.5			FTH1-Z	CGACAGGTAAACGTAGCAGCC	250	
GIF		0.33-0.50	b	214	GIF-A	CCGCAATTCCACAACCC	166	60
		11p11.2-q13.1			GIF-Z	GTTGAACAGCAGCTCAACCC	166	
GSTP1	GST3	0.48-0.59	c	313	GST(1)	CCTGCAGATCTCCTTCCCTGACTACAACCT	833	65
		11q13.1-q13.5			GST(2)	AGAACCCCTAAGCCCGTCCCAGGAAACAC	833	

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
H19		0.00-0.05 11p15.5-p15.4	b	440	H19-A H19-Z	TTGTCAGTAGAGTGCGCCC ACCTCCCTCAGGGTCCAG	833 833	56
HBB		0.00-0.03 11p15.5	c	536	KM29 RS42	GGTTGCCCAATCTACTCCCACG GCTCACTCAGTGTGGCAAAG	833 833	56
HPX		0.00-0.05 11p15.5-p15.4	c	454	HPX-A HPX-Z	AGCTCAGCATGGCTAGGGTA GTAGCATCAAAGCTCCAGCC	500 500	58
HRAS		0.01-0.03 11p15.5	abc	312	H3' H5'	CGGTGCTGAGACGAGGCACT GAGACCCTGTAGCAGGACCC	500 500	56
IL-1BCE		0.78-0.91 11q23.1-q23.3	c	157	IL-1BCE-A IL-1BCE-Z	TTCATTTGAGCAGCCAGATG CCAAAAACCTTTTACAGAACGA	500 500	60
INS		0.01-0.03 11p15.5	abc	439	INS-A INS-Z	CCCTCATTTGATGACCCC GGGACACAGCAGGACACACT	500 500	58
INSL2		0.31-0.59 11p11.2-q13.5	c	507	INSL2-A INSL2-Z	ACTGAGGCTTTTGCAAGGAA CATGCATGAATGAGCATTCC	833 833	56
KRN1		0.48-0.59 11q13.1-q13.5	c	511	KRN1-A KRN1-Z	CTTGTGGCTGCTCCCAGT ACATATGGGGGCATGTGG	500 500	62
LDHA		0.10-0.23 11p15.1-p14.1	bc	523	LDHA-A LDHA-Z	TAGTGTTCCTTGCATTTTGGG ATCCCAGGATGTGACTCACTG	500 500	56
LDHC		0.05-0.19 11p15.4-p14.3	bc	150	LDHC-A LDHC-Z	ATTGTCACAGCAGGTGCAAG AGACTCACCTGGATTTGAAACA	500 500	56
MDU1		0.42-0.50 11q12.1-q13.1	bc	182	MDU1-A MDU1-Z	GAAACTGGAGCCTCACGAAG AACACCCTATTTGGGGGTTC	833 833	56

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (11 different)	FLplex range band(s)	Map loch.	Product size (bp)	Primer name	Primer sequence	(primer) (nM)	Annealing temp.(°C)
MLP		0.00-0.05 11p15.5-p15.4	b	870	MLP-A MLP-Z	ACAACTACCACTGCCATCCC ACTGGTACACCTCCAACCCC	166 166	56
MUC2		0.00-0.03 11p15.5	bc	155	MUC2-B MUC2-Y	CATTCTCAACCACAACCCCT CCAACGATGTTAGTGCCC	250 250	58
MYOD1		0.05-0.13 11p15.4-p15.1	bc	210	MYOD1-B MYOD1-Y	TCGCAGACCTAACCCTGC CAGAGTAAATACAGCCCCAG	500 500	55
NCAM		0.82-0.82 11q23.1	abc	470	JHE312 JHE311	GGCCCCCGTAGAAGCAAACCCAGTGCCA CGTTCTGAGCTTCCTAGATTCTTAATTTC	500 500	66
NFRKB	NF-RKB	0.94-1.00 11q25	c	470	6380 6381	GTACATGCAGCTGATAGCCCTCCCAAGCCC CACGTGCACACACCGTGGTGCTGACAGGG	833 833	60
NGK2		0.00-0.13 11p15.5-p15.1	a	176	NGK2-A NGK2-Z	ATCATCTTCCTGGCCTTGG TACATGTCTCCATAGCCCAGG	500 500	56
OSBP		0.40-0.42 11q11	bc	156	OSBP-A OSBP-Z	ATAGGGACAATACTGGGAGTG GCACTTGGTAAGAGGTCACAA	250 250	57
PBGD		0.87-0.89 11q23.3	abc	406	PBGD-A PBGD-Z	CAAAGGAGCGCCATAGAAG CTGAGGATGCCAACCTGG	250 250	53
PC		0.40-0.50 11q11-q13.1	bc	125	PC-B PC-Y	CTCAGTGCCATGAAGATGGA CAAGATCACTCGATCTCCAGC	250 250	56
PGAJ		0.48-0.59 11q13.1-q13.5	a c	156	PGA(1) PGA(2/2)	CTCATCAGAAAGAAGTCCTTCAGCCCACC ATCCAGGTAGTTCTCCAGGGCTGTTCATC	833 833	52

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range bands(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	(primer) (nM)	Annealing temp.(°C)
PGA5		0.48-0.59 11q13.1-q13.5	c	432	PGA5-A PGA5-Z	TGAGAAAGTACCCTGAGAGC ACCCCTCTAGTGGATGGTCC	500 500	50
PGR		0.70-0.78 11q22.1-q22.3	bc	504	PGR-A PGR-Z	TCCCACACAACATCCAAACT CAACCCCTCCTCCTCTAGG	250 250	56
PPP1CA	PPP1A	0.48-0.59 11q13.1-q13.5	c	208	PP1(1/2) PP1(2/2)	GAACTACGGCGAGTTCAGTGCCCT ATACACCAAGGTCCATGTTCCCCGT	833 833	60
PROS30		0.05-0.24 11p15.4-p13	b	165	PROS30-A PROS30-Z	TCCTCAACCTCCCTGATGAAC TAGCATTCCACCACTCTCCA	250 250	54
PTH		0.09-0.12 11p15.2-p15.1	abc	528	PTH-A PTH-Z	CTCCTCAAAACCAACCCAAT TCACATCAGCTTTCCTCCC	833 833	60
PYGM		0.47-0.49 11q12.3-q13.1	abc	431	PYGM-A PYGM-Z	CTCCTCTCCTCCCCTTCC TCTCATAGTAGTCCTCCTCCC	250 250	56
RAG1		0.24-0.33 11p13-p11.2	b	237	RAG1-D RAG1-Y	ATCAGCACAACTTTCAGAAAATCA TGGAAACTCTAAATTTCTTCAAATC	833 833	52
RBTNL1		0.23-0.27 11p13	c	152	RBTNL1-A RBTNL1-Z	ATCCCCCACCCCTACTTC TACCTCAGATAGTCTCTCCCCC	833 833	56
RNAHX		0.83-0.91 11q23.3	c	175	RNAHX-A RNAHX-Z	CTGTCACACATCCTTTTCCC ACCACACAGTCCAAACAAGTC	833 833	58
RNH		0.00-0.03 11p15.5	c	141	RAI-A RAI-Z	TTCCTCCTCCTCCTCCTCC TCTCCAAAATATACTCCCACAA	250 250	57
RRM1		0.00-0.05 11p15.5-p15.4	c	250	RRM1-A RRM1-Z	TTCCTTCAGGTCCTAACCCT TCCTAAATCCCTCATCAAACC	833 833	54

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FISH range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	[primer] (nM)	Annealing temp.(°C)
SAA		0.10-0.23 11p15.1-p14.1	bc	469	SAA-B	CACCTGGGAATACCAGTGCT	500	48
					SAA-Y	AAGGAACGAAAGAAGCTTCG	500	
SMPD1		0.03-0.05 11p15.4	bc	245	SMPD1-B	AAGAACCAGTCCCTGGCC	250	56
					SMPD1-Y	TCGCAGCCAGCAGTACAGG	250	
SPI1		0.27-0.31 11p12	c	218	SPI1-A	ACCAGTTCCTGTTCGACCTG	500	56
					SPI1-Z	CTTCACCTTCCTTGACCTCC	500	
SRPR		0.91-0.92 11q24.1	abc	150	SRPR-A	GTGGCATGTTTGGTATGCTG	500	56
					SRPR-Z	TGGCAACAGATTCACAGAGC	500	
SSRP1		0.33-0.50 11p11.2-q13.1	b	162	SSRP1-A	TGCCATCTGAACATGCTCTC	250	57
					SSRP1-Z	GGTCGGGAAGTAAGATGAGG	250	
STMY1		0.75-0.78 11q22.3	abc	270	STMY1-A	TGTAGAAGGCACAATATGGGC	833	56
					STMY1-Z	GCAAGCTAAGCAGCAGCC	833	
STMY2		0.75-0.70 11q22.3-q21	bc	185	STMY2-A	GGGGAAGACAGATATGGGT	833	56
					STMY2-Z	CTGTTCAGTGCAATTCAAAGCC	833	
TCN1		0.40-0.48 11q11-q12.3	bc	161	TCN1-B	TGGTAGTTACGTTGTCCCCA	833	56
					TCN1-Y	TCTCCTCCTCGTAGTGGCAT	833	
TH		0.03-0.03 11p15.5	bc	230	TH-A	CCTGGTCCTGCACTGTCC	250	61
					TH-Z	CAGCAGACAGTGTCAGGGAA	250	
TPH		0.10-0.19 11p15.1-p14.3	bc	153	TPH-A	GCGTCCATTTGCAGTGAAGT	833	57
					TPH-Z	TAGATAGTCGGCTTCCTGTG	833	
TTG2		0.24-0.33 11p13-p11.2	b	160	TTG2-A	TTGTTGTGACTTTGAGCCTC	833	58
					TTG2-Z	TTAGGCCTTGGGAAGG	833	

STS MAP OF CHROMOSOME 11

Locus Name	STS Name (if different)	FLpter range band(s)	Map tech.	Product size (bp)	Primer name	Primer sequence	(primer) (nM)	Annealing temp.(°C)
TYR		0.68-0.68 11q21	abc	507	TYR-A	GTCACTCCAATTAGCCCAGTTCC	833	56
					TYR-Z	TCCCGATGACATAGTCTGACC	833	
D11S1003	ZNF1	0.83-0.88 11q23.2-q23.3	a	195	ZNF1-A	TCAAGCCATACACCATTGTCA	250	50
					ZNF1-Z	AATACACCGAATCCCACCAG	250	
D11S1007	ZNF6	0.55-0.56 11q13.3-q13.4	ab	178	ZNF6-A	TTCCGTATCTCCTCCTGCC	833	58
					ZNF6-Z	TCCCCTCTCTTCCCCAAC	833	

a Locations of loci were were determined by FISH (a), cell hybrid analysis (b), and from the GDB (c). Those without regional mapping information are shown as nd.

Data Analysis

As described above, 371 DNA sequences fragments, determined at one pass accuracy, comprising 116 kb of chromosome 11 derived from cosmid ends were analyzed for the presence of repetitive sequences, simple sequence repeats and similarities to known genes. All of the sequence fragments were subjected to computer analyses for the presence of noteworthy sequence structures (Table 7). The presence of repetitive sequences was determined using the program FASTA and a repetitive sequence database (Jurka et al., 1992, J. Mol. Evol. 35, 286-291) supplemented with a comprehensive set of di- and tri-nucleotide repeats. A FASTA cutoff score of 100 determined visually was used to recognize repetitive sequences from background random matches. Similarities to known genes were identified with the program BLAST and the GenBank database.

Amino acid comparisons were performed by translating DNA sequence fragments into all six potential reading frames and comparing translations to protein sequences in the Swiss-Prot, GenPept or PIR databases using the program BLASTX. Putative exons were identified using the program GRAIL on the Oak Ridge National Laboratory Internet server. The results of these various searches were evaluated numerically and by inspection. The data associated with this project, including DNA sequence file pointers, predicted STS primer sequences, test and mapping results, and *in situ* hybridization analysis, were stored in a relational database called Genome Notebook specifically designed for this project. DNA sequence and mapping information on other genes was imported into the Genome Notebook database from GDB and GenBank.

STS MAP OF CHROMOSOME 11

TABLE 7

Analysis of 371 random cosmid end sequences determined by automated fluorescent sequencing

Category subcategory	number	percentage total	percentage category
Contain repetitive DNA	150	40%	
Alu	60	16%	(40%)
LINE-1	41	11%	(27%)
Middle element repetitive	17	5%	(11%)
CA repeats	7	2%	(5%)
others	27	7%	(18%)
Grail predicted exons	34	9%	
Excellent	6	2%	(18%)
Good	9	2%	(26%)
Marginal	19	5%	(55%)
Matches to protein sequences	29	8%	
Certain	7	2%	(24%)
Probable	2	1%	(7%)
Likely	2	1%	(7%)
Possible	8	2%	(28%)
Marginal	10	3%	(35%)

Note: Repetitive sequence analysis was carried out using FASTA and a customized repetitive sequence database. Similarities to known proteins was determined using BLASTX. Putative exons were determined using GRAIL. The frequency of a sequence fragment containing the element is shown in bold and subtypes of repetitive DNA, potential exons and sequence match quality is shown as well as percentage of total. The last column shows the percentage of sequences within each category

The results indicate that the average length of reliable sequence was 312 nucleotides with a standard deviation of 46 nucleotides. Repetitive DNA sequences of some type were found in 150 of these sequences with Alu elements (40%) and LINE-1 sequences (27%) being the most frequent. Middle element repetitive sequences (11%) and simple sequence (CA)n repeats (5%) were also detected with reasonable frequency. The neural net based program, GRAIL, was utilized to predict the locations of possible exons and detect putative genes in 34 sequences (9%); half of these were rated excellent or good according to reliability estimates used by this program. Analysis for additional possible gene sequences was carried out by computer searches for identity or similarity matches at the nucleotide and amino acid level. Significant matches to known protein sequences were detected for twenty-nine (8%) of the sequence fragments (see, e.g., Table 8).

STS MAP OF CHROMOSOME 11

TABLE 8

Results of searching the six-potential coding frames of 371 cosmid end sequences against the protein sequence databases with BLAST-X

Likely significance sequence name	map position	potential homologue
Certain		
D11S384[a]	11q23.1	mitochondrial acetoacetyl-CoA thiolase
c11q-2b11-t	11q23.3	UDP-N-acetylglucosamine-dolichyl-phosphate N-acetylglucosaminephosphotransferase
cSRL-2e4-t[b]	11p15.4-p13	postsynaptic density protein
cSRL-4a3-t[c]	----	Zinc finger protein ZFP-37
cSRL-5f2-t	11p15.5-p15.4	retrovirus related POL polyprotein
cSRL-7d2-t	----	opioid binding protein/cell adhesion molecule
Probable		
cSRL-2e4-t[b]	11p15.4-p13	discs-large tumor suppressor
cSRL-6g5-t	----	env polyprotein - feline endogenous virus ECE1
Likely		
cSRL-2c2-t	11p11.2-q13.1	mitochondrial carnitine palmitoyltransferase precursor
cSRL-2g6-t	11p15.4-p13	GATA-3 transcription factor

Likely significance sequence name	map position	potential homologue
Possible		
cSRL-2f5-t	11p15.4-p13	biglycan
cSRL-3a3-t	11p13-p11.2	cytochrome P450 IIA6 (Coumarin 7-hydroxylase)
cSRL-3a6-t	-----	potassium channel protein
cSRL-4e2-t	-----	phosphomannomutase
cSRL-4h2-t	-----	properdin, mouse
cSRL-5b11-t	-----	hypothetical 137.7 KD protein in subtelomeric Y' repeat
cSRL-6b4-t	-----	Wnt-4 protein
cSRL-7e6-t	-----	olfactory receptor protein

Note: Matches are grouped into several categories of likely significance from certain to marginal.

a The D11s384 sequence was determined manually and is not part of the 371 sequences determined with automated techniques

b Two significant matches, from the same gene family, were found to the same cosmid.

c This cosmid was derived from hamster based on PCR amplification results.

The alignments of nine of the best matches are shown in Figure 11 including matches to mitochondrial acetoacetyl-CoA thiolase, UDP-N-acetylglucosamine-dolichyl-phosphate N-acetylglucosamine phospho-transferase, postsynaptic density protein (and the related discs-large tumor suppressor), zinc finger protein ZFP-37, retrovirus related pol polyprotein, opioid binding protein/cell adhesion molecule, the env polyprotein from feline endogenous virus ECE1, mitochondrial carnitine palmitoyltransferase II precursor, and the trans-acting T-cell specific transcription factor GATA-3. The zinc finger protein match is probably not of a human genomic DNA origin since PCR amplification with primers predicted from its nucleotide sequence only generate specific products from hamster DNA. One sequence fragment from a cosmid containing the marker D11S384 was determined with manual methods and exactly matched 72 bases of mitochondrial acetoacetyl-CoA thiolase exon six, mapping this gene to 11q23.1 or Flpter 0.78. This gene was recently localized independently to chromosome 11q22.3-q23.1, in agreement with the localization of this sequence, by traditional mapping methods. The detection of a DNA sequence encoding a protein fragment with 97% protein sequence identity (68/70 amino acids) to the hamster UDP-N-acetylglucosamine-dolichyl-phosphate N-acetylglucosaminephosphotransferase strongly suggests that the human homolog of this enzyme, or a highly conserved pseudogene, is located on chromosome 11q23.3. The remaining protein sequence matches range from possible to marginally significant. The significance of some of the detected matches is remarkable and makes gene localization to these regions likely.

The identification of eight putative new genes from 371 cosmid end sequences corresponds to a 2% gene identification rate and the 19 likely or possible genes

corresponds to 5% (Table 8; Figure 11). In comparison, the rate of gene identification from brain cDNA sequencing is about 14%. Strategies for sequencing cDNA libraries suffer from the problem of sequencing the same cDNA multiple times due to the differential abundance of mRNAs. As demonstrated here, random genomic sequencing is associated with a reasonable rate of gene identification (2-5%) coupled with direct gene mapping, and thus is considered to be an advantageous strategy for further characterization of cosmid and YAC clone maps. STSs prepared by the methods described herein, will provide a useful reagent set for further physical mapping, for constructing YAC contigs using STS content mapping, and for further DNA sequence analysis. In addition, automated fluorescent sequencing of randomly chosen cosmid clones is a rapid and powerful tool for generating PCR detectable markers as well as defining the locations of putative genes. Theoretical analysis of the strategy of STS content mapping (Arratia et al., 1991, Genomics, 11:806-827; Palazzolo et al., 1991, Proc. Natl. Acad. Sci., USA, 88:8034-8038) suggests that this number of unique and uniformly distributed markers will serve as an appropriate starting point for future physical analysis.

In summary, sequences were determined from the ends of chromosome 11 specific cosmids by automated sequencing without intermediate subcloning. The STSs and cosmids were mapped by *in situ* hybridization, somatic cell hybrid analysis or both. This effort generated 370 STSs specific for human chromosome 11 and regionally mapped most of them. Sixty-eight percent of these STSs (251/370) were produced from new chromosome 11 sequences; 18% (68/370) represented sequences derived from cloned genes; 8% (29/370) were based upon STS markers deposited and available from GDB. The latter were retested with

our set of standard conditions to allow integration of this map with the results of other groups.

While the invention has been described in detail with reference to certain preferred embodiments thereof, it will be understood that modifications and variations are within the spirit and scope of that which is described and claimed.

Sequence Listing

SEQ ID NO:1
GGTCAAGCTC AGCAACATGA

SEQ ID NO:2
TGCTTTGTGA CCATCGAGAG

SEQ ID NO:3
CAGCAGATGG TCAAGCAAAA

SEQ ID NO:4
ACTCCTGACA CCACCACCTC

SEQ ID NO:5
TCGCTCACTG ACTCGCTG

SEQ ID NO:6
AGCCCTCCCG TATCGTAGTT

SEQ ID NO:7
CTTGAGAGCC TTCAACCCAG

SEQ ID NO:8
AACTGGGCGG AGTTAGGG

That which is claimed is:

1. A method for sequencing complex genomes, said method comprising:

(1) sequencing the end-specific nucleotides of each member of a library of cosmid clones,

wherein said cosmid clones are prepared by inserting genomic DNA fragments into cosmid vectors, and

wherein the cosmid vectors include sequences of nucleotides that flank at least one end of the inserted DNA, and that serve as transcription initiation sites for the synthesis of nucleic acids specific to the ends of the inserted DNA, and

(2) assembling a sequence sampled map by correlating the end-specific nucleotide sequence information with the relative spatial relationship between the cosmids.

2. A method according to Claim 1 wherein the relative spatial relationship between the cosmids has been determined prior to said sequencing the end-specific nucleotides of each member of a library of cosmid clones.

3. A method according to Claim 1 wherein the relative spatial relationship between the cosmids is determined by the cosmid multiplex analysis method.

4. A method according to Claim 1 wherein the relative spatial relationship between the cosmids is determined by restriction-fragment-length mapping of the cosmids.

5. A method according to Claim 1 wherein at least 100 base pairs of end-specific nucleotide sequences are determined.

6. A method according to Claim 1 wherein said cosmid clones are generated in cosmid vectors allowing for the synthesis of end-specific nucleic acid sequences directly from at least one end of DNA fragments inserted therein.

7. A method according to Claim 6 wherein said cosmid vectors comprise at least one promoter specific for a bacteriophage RNA polymerase and a cloning site allowing for the insertion of DNA fragments, said promoter being positioned operatively for transcription of a DNA fragment inserted into said cloning site.

8. A method according to Claim 7 wherein said cosmid vectors comprise two oppositely oriented promoters, each of which is specific for a bacteriophage RNA polymerase, positioned on two sides of said cloning site, operatively for transcription of a DNA fragment inserted into said cloning site.

9. A method according to Claim 8 wherein each of said bacteriophage RNA polymerase-specific promoters is selected from the group consisting of promoters specific for bacteriophage T7 RNA polymerase, and promoters specific for bacteriophage T3 RNA polymerase.

10. A method according to Claim 9 wherein said cosmid vector is selected from the group consisting of pWE8, pWE10, pWE15, and pWE16.

11. A method according to Claim 6 wherein said cosmid vectors comprise at least two cos sites.

12. A method according to Claim 11 wherein said cos sites are separated by unique restriction sites.

13. A method according to Claim 12 wherein said cosmid vector is selected from the group consisting of sCOS-1, sCOS-2, sCOS-4, and derivatives thereof.

14. A method for sequencing complex genomes, said method comprising:

(1) preparing a genomic library of cosmid clones by inserting DNA fragments from said genome into cosmid vectors, wherein the cosmid vectors include sequences of nucleotides that flank at least one end of the inserted DNA, and that serve as transcription initiation sites for the synthesis of end-specific probes,

(2) arranging the cosmid clones, whereby each clone may be identified and replicas of said arrangement may be reproduced,

(3) pooling portions of said cosmid clones and synthesizing pools of mixed end-specific probes from the DNA inserts that have been prepared from said pooled clones, wherein each pool contains fewer than all of the cosmid clones in the library, but all of the cosmid clones in the library are included in at least one pool,

(4) hybridizing each pool of probes to a replica of said arranged cosmid clones and identifying the cosmid clones in each replica that hybridize to the probes, wherein said identified clones include the pooled cosmid clones and cosmid clones that contain DNA inserts that overlap with the DNA inserts in the pooled clones,

(5) identifying the cosmid clones from among those identified in step (4) that hybridize to two or more pools of probes, thereby identifying groups of cosmid clones that include overlapping DNA,

(6) assembling contigs from said groups, and

(7) sequencing the fragment ends of the DNA inserts of each of the overlapping cosmid clones.

15. A method according to Claim 14 wherein cross-hybridizing clones are identified by comparing the data sets obtained from two groups of cosmid clones containing at least one common clone, and repeating the pairwise comparison with other groups of clones containing at least one common clone.

16. A method according to Claim 14 wherein said cosmid clones are pooled according to the rows and columns of a two-dimensional matrix, and said mixed end-specific RNA sequences are hybridized to a replica of the entire matrix.

17. A method according to Claim 14 wherein said cosmid clones are pooled according to the planes intersecting with a three-dimensional matrix, and said mixed end-specific RNA sequences are hybridized to a replica of the entire matrix.

18. A method according to Claim 14 wherein said cosmid clones are generated in cosmid vectors allowing for the synthesis of end-specific RNA sequences directly from at least one end of DNA fragments inserted therein.

19. A method according to Claim 18 wherein said cosmid vectors comprise at least one promoter specific for a bacteriophage RNA polymerase and a cloning site allowing for the insertion of DNA fragments, said promoter being positioned operatively for transcription of a DNA fragment inserted into said cloning site.

20. A method according to Claim 14 wherein said cosmid vectors comprise two oppositely oriented promoters, each of which is specific for a bacteriophage RNA polymerase, positioned on two sides of said cloning site, operatively for transcription of a DNA fragment inserted into said cloning site.

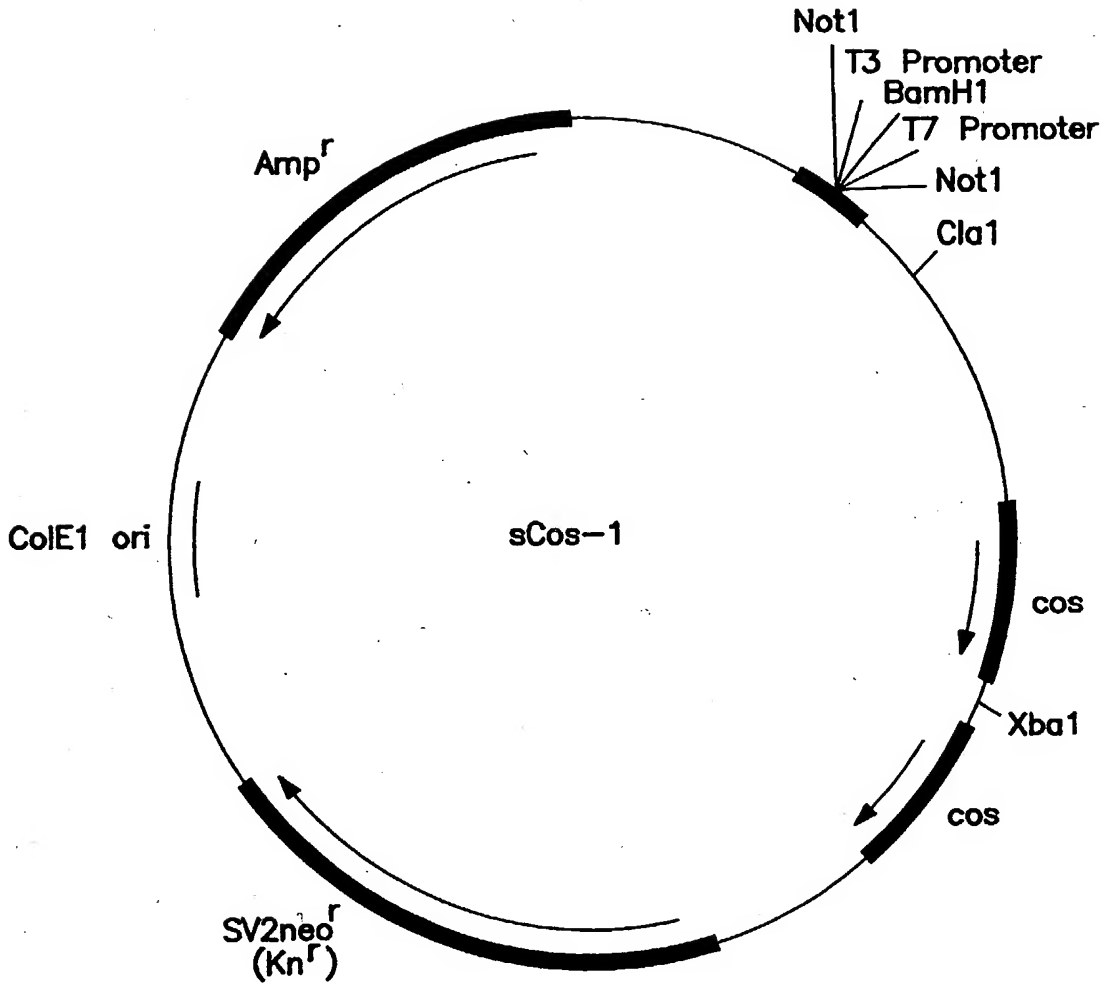
Not1
T3 Promoter
BamH1
T7 Promoter
Not1
Cla1
Amp^r
ColE1 ori
sCos-1
cos
Xba1
cos
SV2neo^r
(Kn^r)

FIG. 1

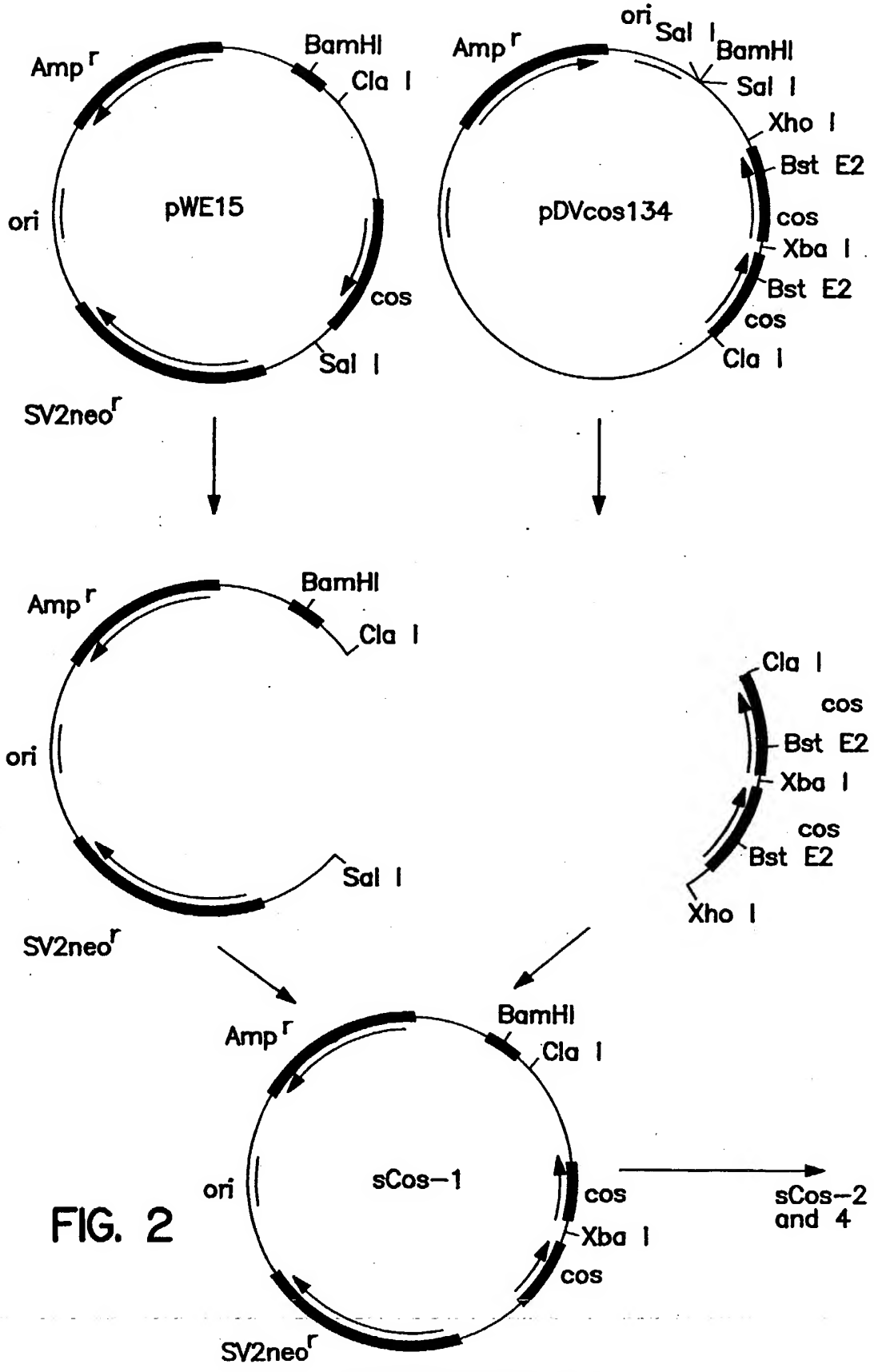
Amp r
BamHI
Cla I
ori
pWE15
cos
Sal I
SV2neo r
ori
Sal I
BamHI
Sal I
Amp r
Xho I
Bst E2
pDVcos134
cos
Xba I
Bst E2
cos
Cla I
Amp r
BamHI
Cla I
ori
Sal I
SV2neo r
Cla I
cos
Bst E2
Xba I
cos
Bst E2
Xho I
Amp r
BamHI
Cla I
ori
sCos–1
cos
Xba I
cos
SV2neo r
sCos–2
and 4

FIG. 2

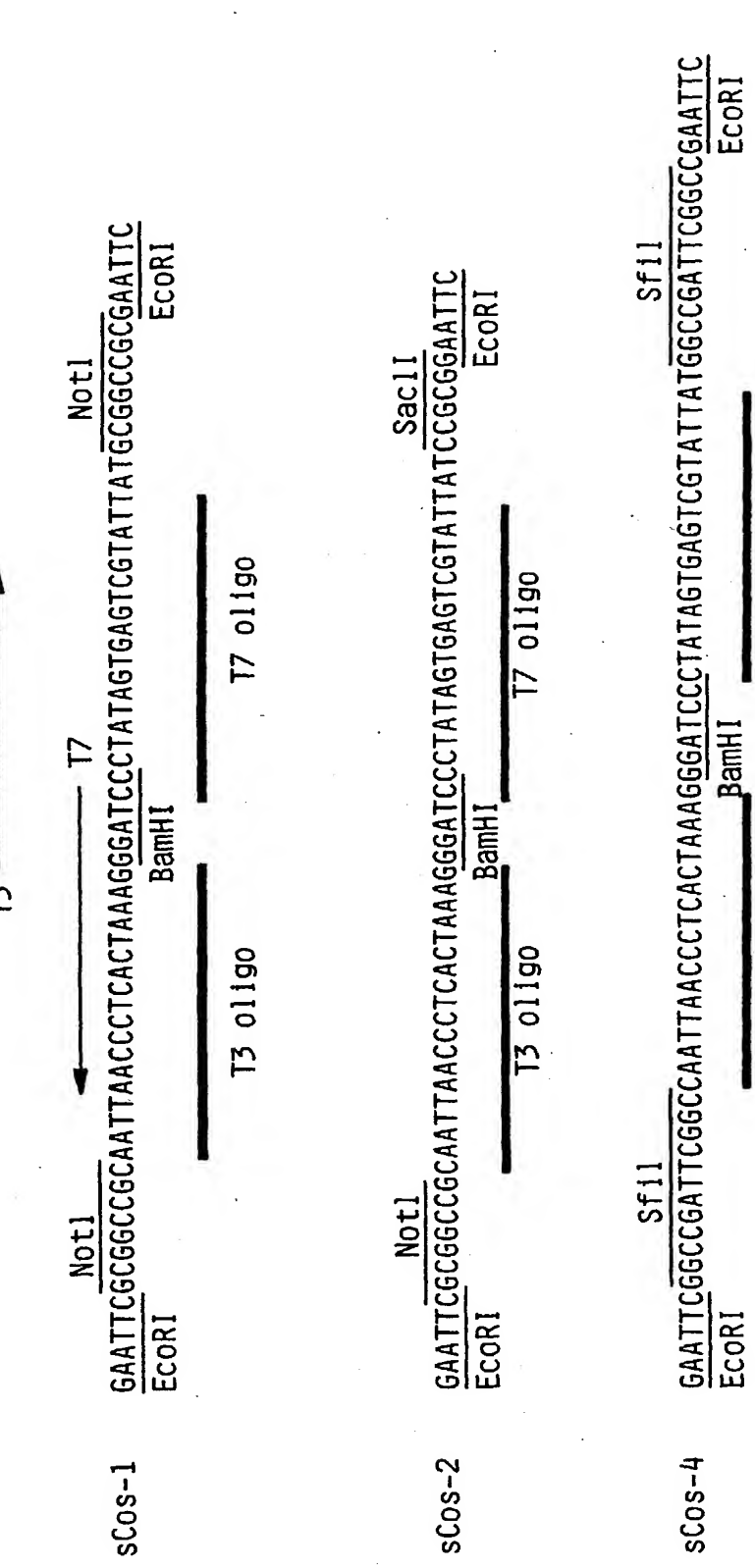
sCos-1
GAATTCGCGGCCGCAATTAACCCTCACTAAAGGGATCCCTATAGTGAGTCGTATTATGCGGCCGCGAATTC
EcoRI
NotI
BamHI
T7
NotI
EcoRI
T3
T3 oligo
T7 oligo
sCos-2
GAATTCGCGGCCGCAATTAACCCTCACTAAAGGGATCCCTATAGTGAGTCGTATTATCCGCGGAATTC
EcoRI
NotI
BamHI
SacII
EcoRI
T3 oligo
T7 oligo
sCos-4
GAATTCGGCCGATTCGGCCAATTAACCCTCACTAAAGGGATCCCTATAGTGAGTCGTATTATGGCCGATTCGGCCGAATTC
EcoRI
SfiI
BamHI
SfiI
EcoRI
T3 oligo
T7 oligo

FIG. 3

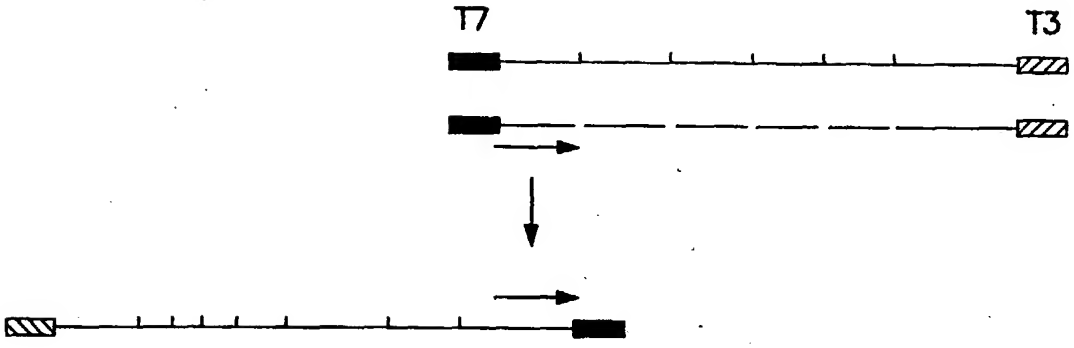

FIG. 4A

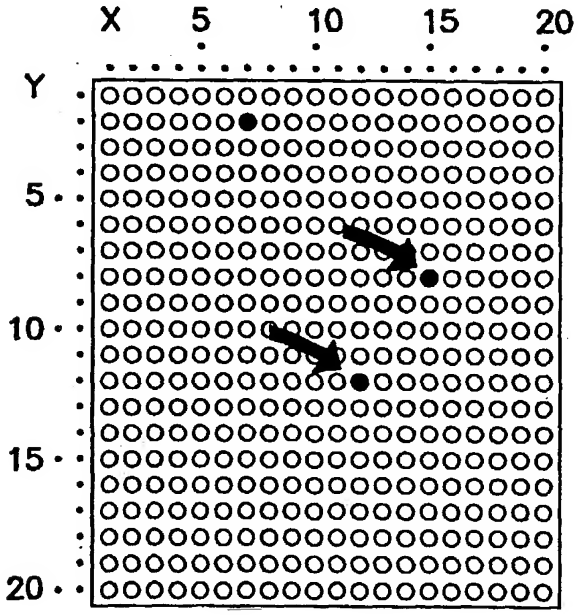

FIG. 4B

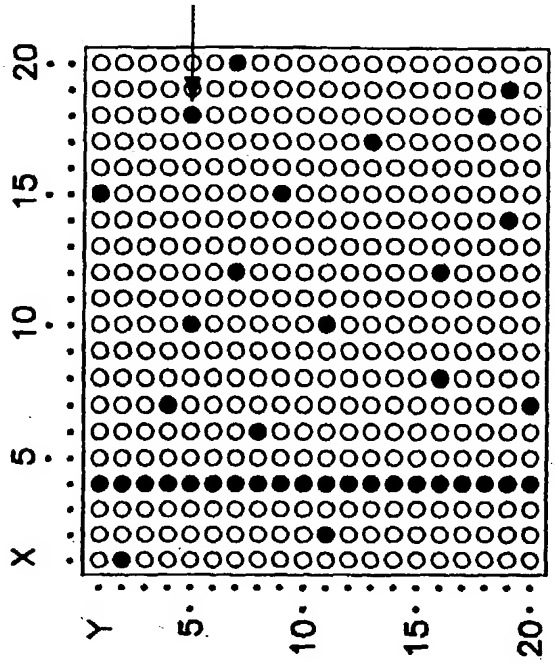

FIG. 4D

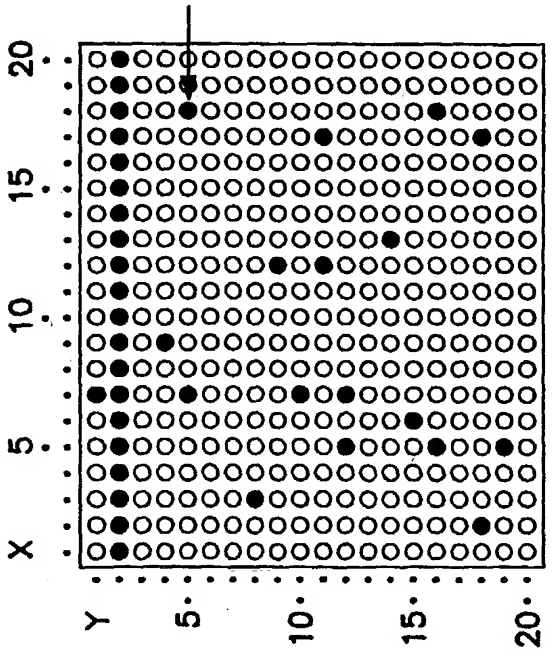

FIG. 4C

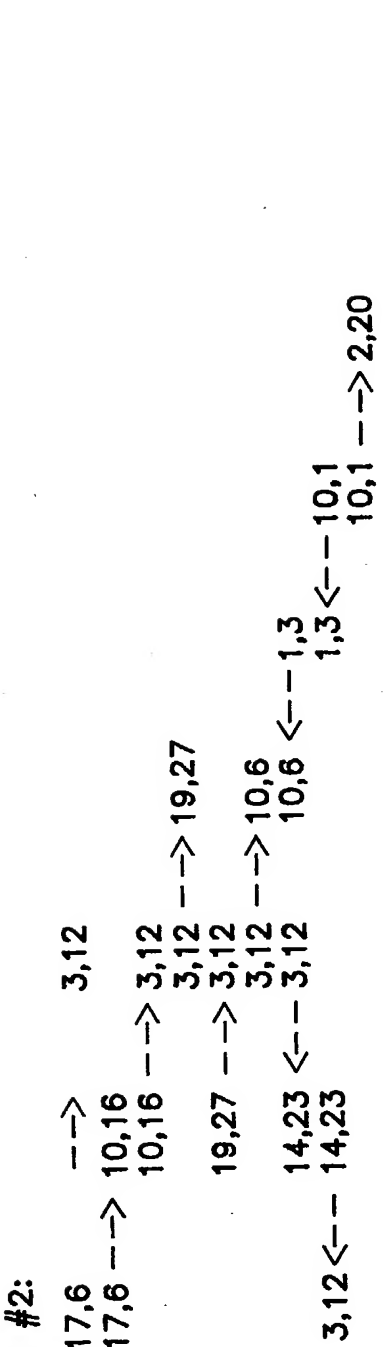
Contig #2:
17,6 --> 3,12
17,6 --> 10,16
10,16 --> 3,12
3,12 --> 19,27
19,27 --> 3,12
3,12 --> 10,6
14,23 <-- 3,12
10,6 <-- 1,3
3,12 <-- 14,23
1,3 <-- 10,1
10,1 --> 2,20

FIG. 5A

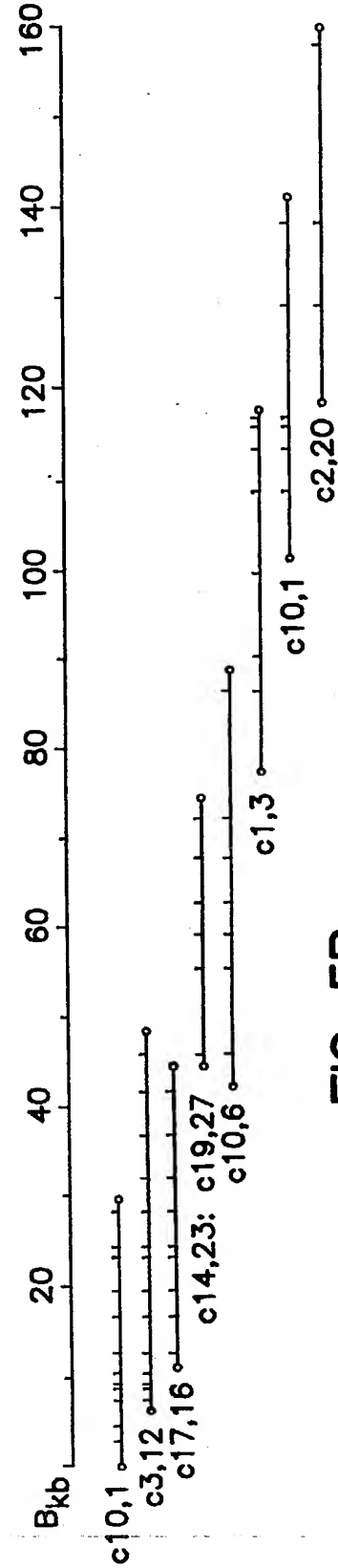
Bkb
20
40
60
80
100
120
140
160
c10,1
c3,12
c17,16
c14,23: c19,27
c10,6
c1,3
c10,1
c2,20

FIG. 5B

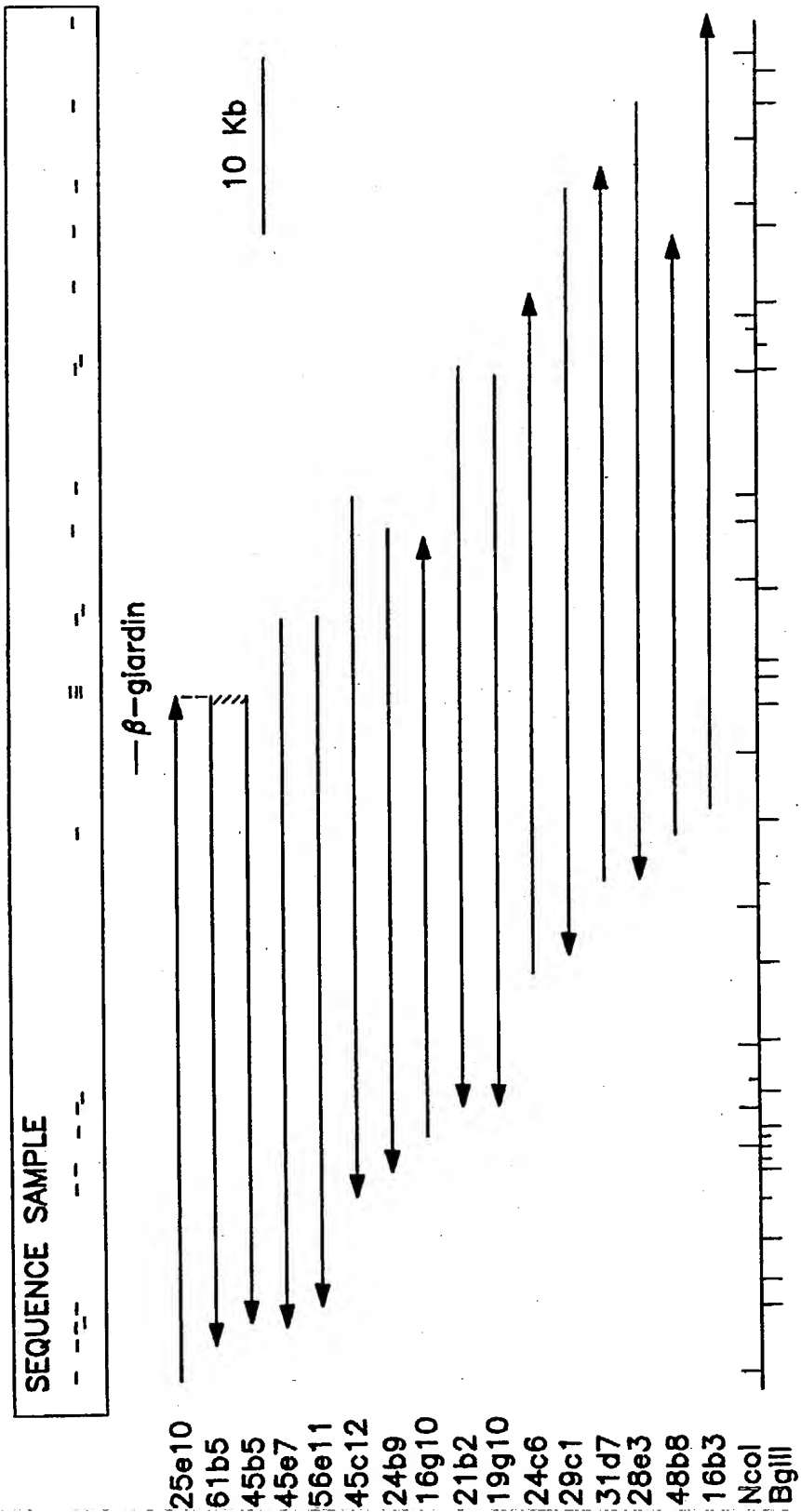
SEQUENCE SAMPLE
10 Kb
—β—giardin
25e10
61b5
45b5
45e7
56e11
45c12
24b9
16g10
21b2
19g10
24c6
29c1
31d7
28e3
48b8
16b3
NcoI
BglII

FIG. 6A

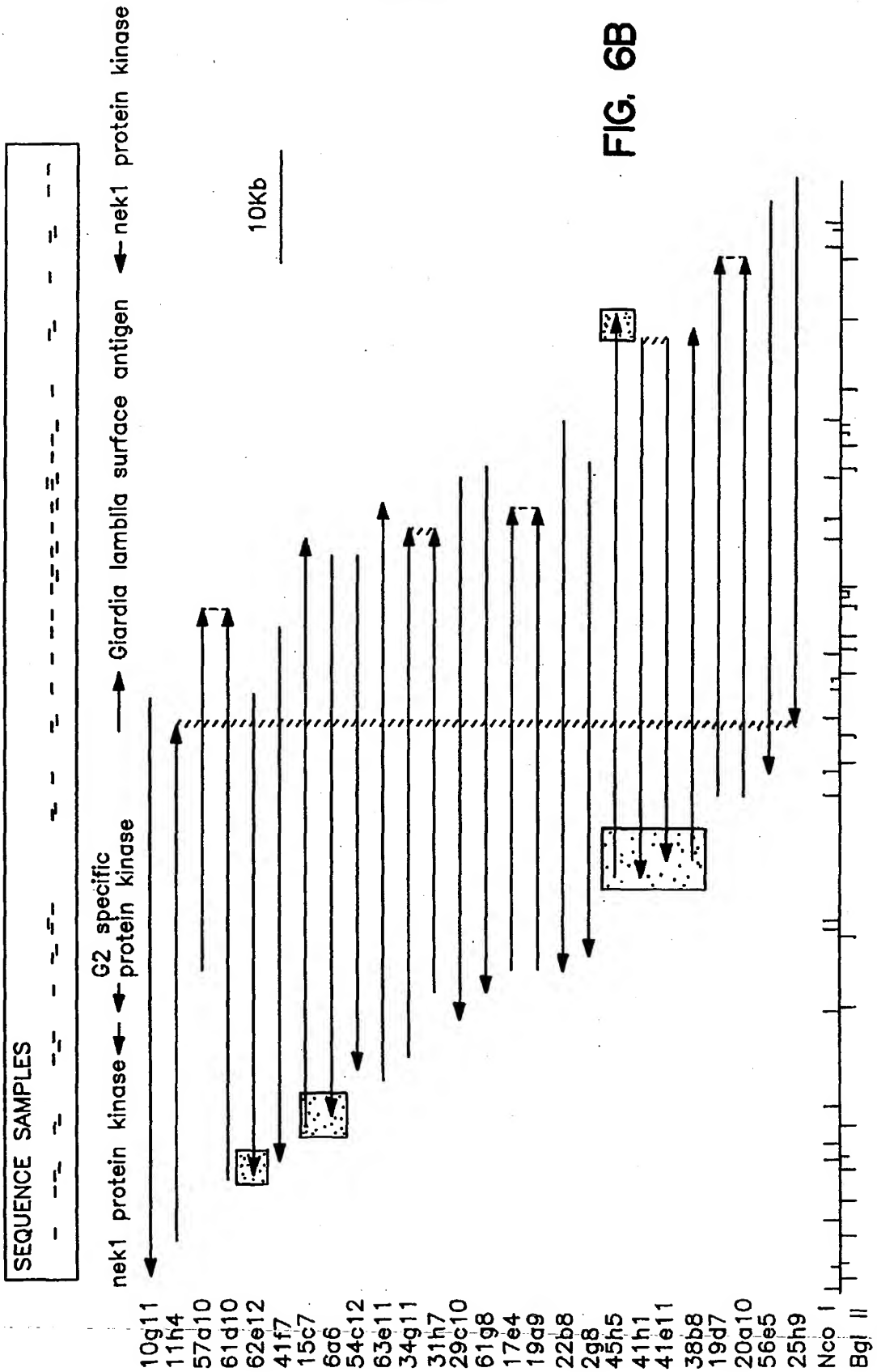
SEQUENCE SAMPLES
nek1 protein kinase
G2 specific protein kinase
Giardia lambila surface antigen
nek1 protein kinase
10Kb
10g11
11h4
57d10
61d10
62e12
41f7
15c7
6a6
54c12
63e11
34g11
31h7
29c10
61g8
17e4
19a9
22b8
2g8
45h5
41h1
41e11
38b8
19d7
20a10
56e5
25h9
Nco I
Bgl II

FIG. 6B

cGR-62e12-u(GLNK)

```
                  * *  *          *     *                  **   *    *** *          *
CDK3      MDMFQKVEKIGEGTYGVVYKAKNRETGQLVALKKIRLDLEMEGVPSTAIREISLLKELKHPNIVRLLDVVHNERKLYL..
CDC2      MEDYIKIEKIGEGTYGVVYKGRHRVTGQIVAMKKIRLESEEEGVPSTAIREISLLKELRHPNIVSLQDVLMQDSRLYL..
KKIA     MMEKYEKIGKIGEGSYGVVFKCRNRDTGQIVAIKKFLESEDDPVIKKIALREIRMLKQLKHPNLVNLLEVFRRKRRLHL..
GLNEK       MYIKNRILGRGAYGIAWLAKDTETGASVVIKELTLAQLPAAERERALREANLLSQLFHPNIVSYKQSFLENGALNT..
NEK1      MEKYVRLQKIGEGSFGKAVLVKSTEDGRHYVIKEINISRMSDKERQESRREVAVLANMKHPNIVQYKESFEENGSLYI..
```

FIG. 7A

cGR-6a6-u and cGR-15c7-t (GLNIMA)

```
          **          *       *     *                  * * * * *  *              ***     *
NIMA     ..FG       IIRKVKRKSDGFILCRKEINYIKMSTKEREQLTAEFNILSSLRHPNIVAYYHREHLKASQDLYLYMEYCGGGDL
KIN3     ..FG       SVRKVIHIPTKKLLVRKDIKYGHMNSKERQQLIAECSILSQLKHENIVEFYNWDFDEQKEVLYLYMEYCSRGDL
GLNIMA  ..FGTRGSGTVTGPSVQZWKSRARLCSQGSQYYKYTGQVQICVRNDFIKLCSLSHKNLVKYYDVYYNDTKNLGHFVMEYYERCCL

          *          *  ** **  *  **    *                                ** **   **  *
NIMA     SMVIKNLKRTNKYAEEDFVWRILSQLVTALYRCHYGTDPAEVGSNLLGPAPKPSGLKGKQAQMTILHRDLKPENIFLGSDNT..
KIN3     SQMIKHYKQEHKYIPEKIVWGILAQLLTALYKCHYGVELPTL TTIYDRM KPP VKGKN   IVIHRDLKPGNIFLSYDDS..
GLNIMA  MDVILFYRMKERVIPEETVWYILSHLAEALL  YY  HSPQKDNTDMGP               LVHRNIKPSKVFLAADGY..
```

FIG. 7B

•repetitive DNA similar to 3' noncoding region surace antigen gene, G. lamblia

```
cGR-38b8-t:    9  CATCTGTNACTCACCCCTCTGTGTGCTCGTGGTCTGTGAGTGCCTCCACCCAGACAACGC
                  ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:       677  CATCTGTGACTCACCCCTCTGTGTGCTCGTGGTCTGTGAGTGCCTCCACCCAGACAACGC

cGR-38b8-t:       CTCAGGAGCCGTGCACACCGGGACAGAGAGGATGACCAAGGGGGTCAAGCAGCCCCGCTA
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:            CTCAGGAGCCGTGCACACCGGGACAGAGAGGATGACCAAGGGGGTCAAGCAGCCCCGCTA

cGR-38b8-t:       CGAGAGATGAGAGTGCACAGCAACCAGCCTGATC  162
                  ||||||||||||||||||||||||||||||||
Sbjct:            CGAGAGATGAGAGTGCACAGCAACCAGCCTGATC  830
```

FIG. 7C

•surface protein 11, G. lamblia

```
cGR-45h5-t:  242  TGSEPNTCKACSAVINGKXYCSQCNS  319
                  TG     CK C   I+G  YCS+C++
Sbjct:       414  TGQGSGACKTCGLTIDGASYCSECDT  439
```

•major trophozoite antigen tsaa, G. lamblia

```
cGR-41h1-u:  237  TGSEPNTCKACSAVINGKKYCSQCNS  314
                  TG+ + +CK+C+  I+G  YCS+C +
Sbjct:       466  TGTGAGACKTCGLTIDGASYCSECAT  491
```

FIG. 7D

variable surface proteins, G. lamblia similarity to cGR-45h5-t and cGR-41h1-u

```
                                          **          *          * *  ***          * *
cGR-45h5-t+41h1-u   ..QASHMKVESARPHLSVQTHAEGLPCQcVM HRQMPRRGGLVNRAVMAcVT DAGTMF YAVLLIGLML
TSAA                ..GATEGAKKLCKECTAANCKTCDDQGQcQACNDGFYKNGDACSPCHEScKTCSAGTASDCTECPTGKAL
TSP11               ..GTNADNKKACKECTVANCKTCNDQGQcQTCNDGFYKNGDACSPCHEScKTCSAGTASDCTECPTGKAL

                          *  * *   **     ** *   * *  *** *       *                *
cGR-45h5-t          EMARAAQAGDEcEE  VTGSEPNTcKAcSAVINGKKYcSQcNSGG  SQSAPTDGKCTTATTEcSQKQD..
TSAA                KYGDDGTKG TcGEGCTTGTGAGAcKTcGLTIDGASYcSEcATTTEYPQNGVCAPKASRATPTcNDSPI..
TSP11               KYGNDGTKG TcGEGCTTGQGSGAcKTcGLTIDGASYcSEcDTQNEYPQNGICTSTTARTVATcKNSNV..
```

FIG. 7E

variable surface proteins, G. lamblia similarity to cGR-41e11-u

```
                        *  * *                    *    *  **    **** *   **
CGR-41e11-u         ..GNDPXDGVcTSIXA TARXASXCKASGGKXTEcGANYALXSGGcYNTQKLPGSSX
VSP1267             ..GYAPIDGIcTAVAA AGRDVSVCTATGGKCTACTGNYALLSGGcYNTQTLPGKSV
TSAA                ..TEYPQNGVcAPKASRATPTCNDSPIQNGVCGTcADNYFKMNGGcYETVKYPGKTV

                    *    *         * *              *   *            ** *  ** *
CGR-41e11-u         cTAAQNKXXX  QTXANGQSP AGGXF PAcTXXcL  KXAXSKETcTDcLAGYYKG..
VSP1267             cKAVANSNDGKCKTCANGQAPDPATNFCPLcDSTcAECSTKNDADAcTKcFPGYYKT..
TSAA                cLSAPNGGT  CQKAADGYKLDSGTL  TVcSEGc   KECASSTDcTTcLDGYVKS..
```

FIG. 7F

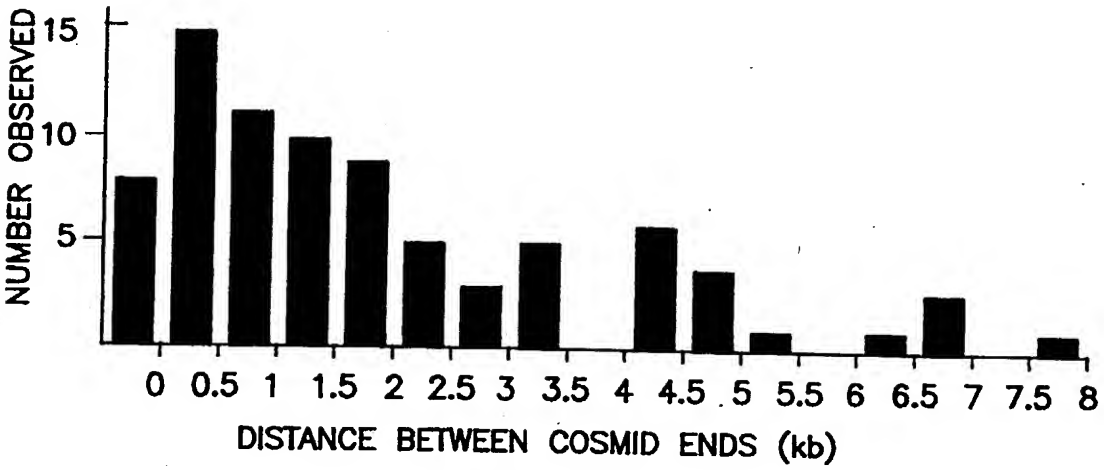
NUMBER OBSERVED
15
10
5
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6 6.5 7 7.5 8
DISTANCE BETWEEN COSMID ENDS (kb)

FIG. 8

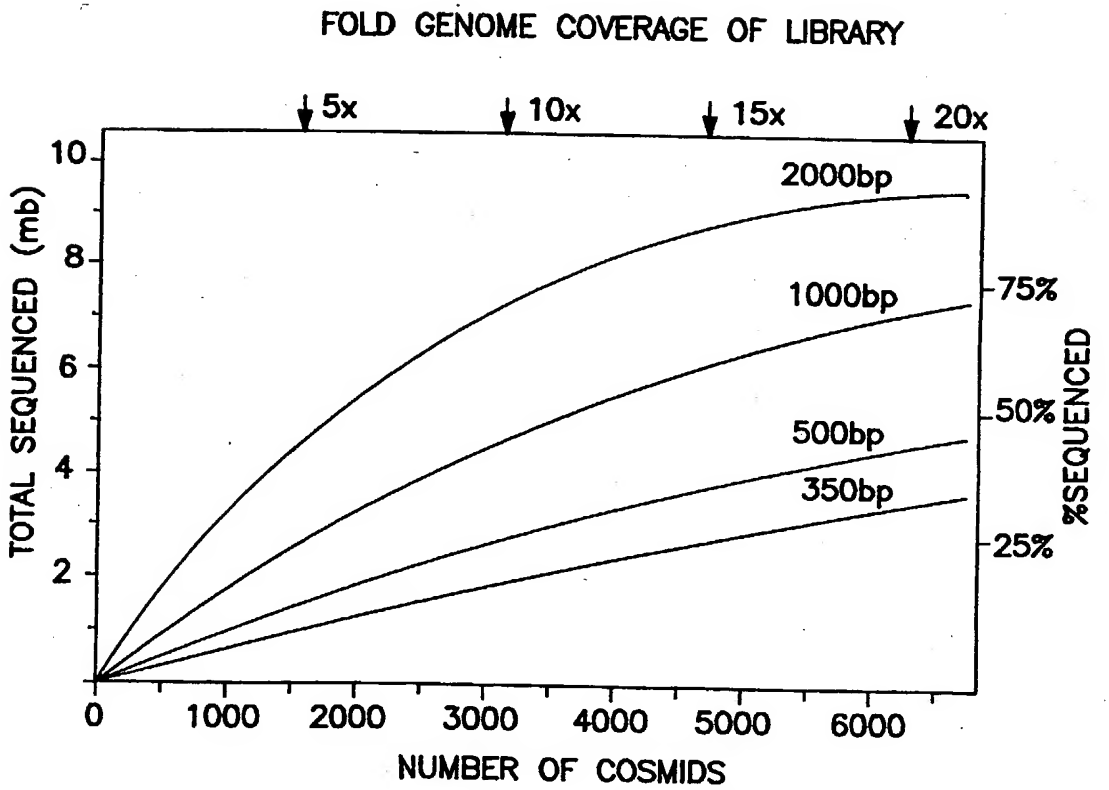
FOLD GENOME COVERAGE OF LIBRARY
5x
10x
15x
20x
TOTAL SEQUENCED (mb)
10
8
6
4
2
2000bp
1000bp
500bp
350bp
75%
50%
25%
%SEQUENCED
0 1000 2000 3000 4000 5000 6000
NUMBER OF COSMIDS

FIG. 9

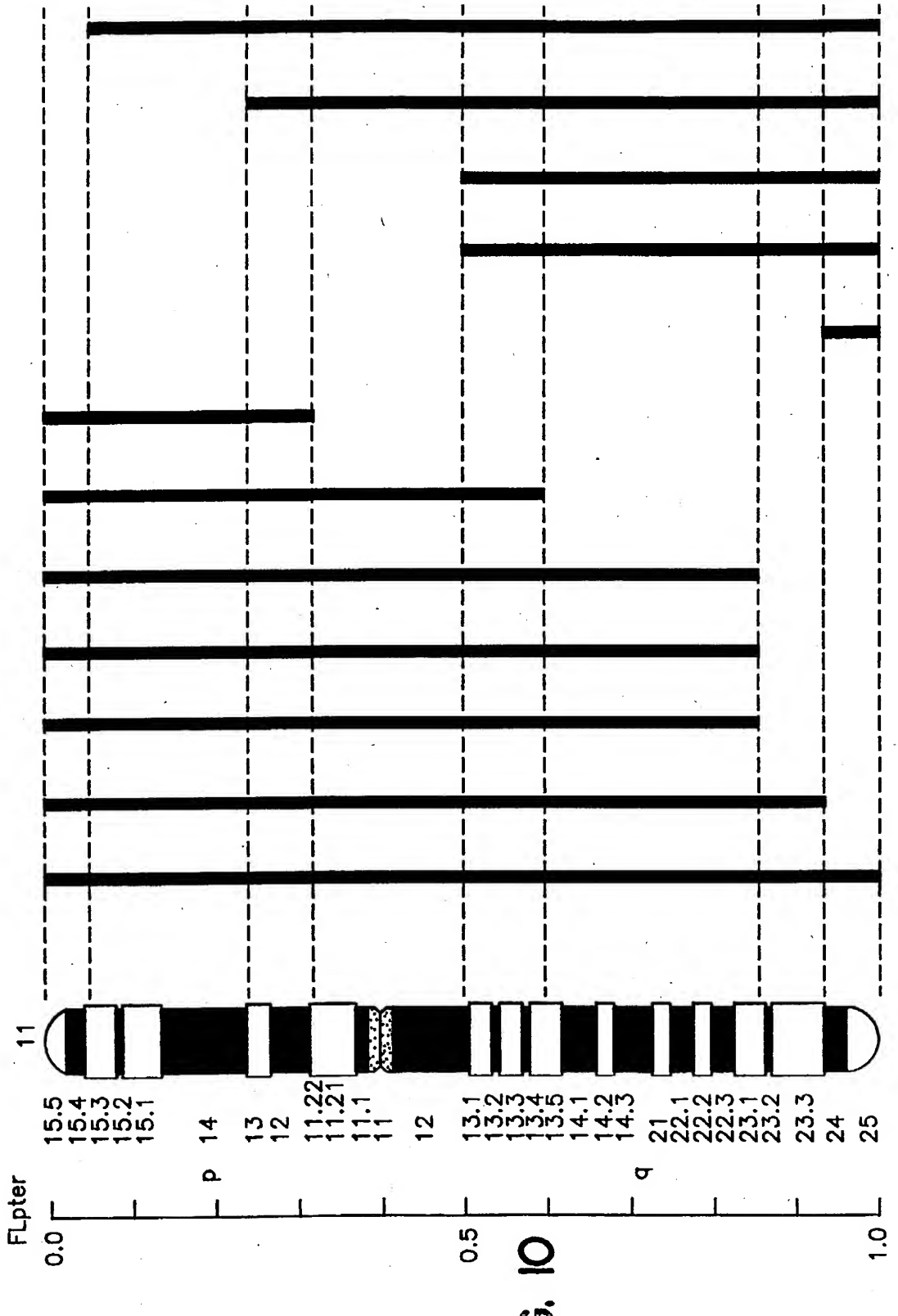
11
15.5
15.4
15.3
15.2
15.1
14
13
12
11.22
11.21
11.1
11
12
13.1
13.2
13.3
13.4
13.5
14.1
14.2
14.3
21
22.1
22.2
22.3
23.1
23.2
23.3
24
25
p
q
FLpter
0.0
0.5
1.0

FIG. 10

mitochondrial acetoacetyl-CoA thiolase, human

```
D11S384        1   KLEDLIVKDGLTDVYNKIHM  60
                   ||||||||||||||||||||
HUMMAT12     174   KLEDLIVKDGLTDVYNKIHM  193
```

FIG. 11A

UDP-N-acetylglucosamine--dolichyl-phosphate N-acetylglucosaminephosphotransferase, hamster

```
c11q-2b11-t  302   ALIGALLAICCMIFLGFADDVLNLRWRHKLLLPTAASLPLLMVYFTNFGKTTIVVPKPFRPILGLHLDLG 93
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||+|||||||||| |||||||||
UAGT_CRIGR    97   ALIGALLAICCMIFLGFADDVLNLRWRHKLLLPTAASLPLLMVYFTNFGNTTIVVPKPFRWILGLHLDLG 166
```

postsynaptic density protein, rat

```
RATPSD95A_1  362      VNGVDLRNASHEQAAIALKNAGQTVTIIAQYKPE  395
                      |||+||| |||||||| ||| ||||||||||| ||
cSRL-2e4-t   148   LFQVNGIDLRGASHEQAAAALKGAGQTVTIIAQYQPE  249
                  |  || ++|  |+|||+|| ||| +|  ||++|||+||
DRCDLGA_1    532   LLSVNNVNLTHATHEEAAQALKTSGGVVTLLAQYRPE  568
```

discs-large tumor suppressor, fruit fly

Zinc-finger protein ZFP-37, mouse

```
cSRL-4a3-t   336   LKVHQRIHTGEKPYQCSDCGKSFTHGSTLKVHQRIHTGXKPYNCNVCGKCFMKGSTLQAH  157
                  |  | |||||||||| |++|||+| |+| |  | | ||| |||+|+ ||| |   |+|  |
ZF37_MOUSE   285   LTDHLRIHTGEKPYKCNECGKTFRHSSNLMQHLRSHTGEKPYECKECGKSFRYNSSLTEH  344
```

retrovirus related POL polyprotein, human

```
cSRL-5f2-t   131   TTTSAEHFTGKKNSPHEGKRIWWKDNKNKTWEIGKVITWGRGFACFSAGENQLPVWXPTR  310
                  ||++ +| |||||||||| |||||||||||||||||||| | |+|||||||| |||
POL1_HUMAN   793   TTSAEQHLTGKKNSPHEGKLIWWKDNKNKTWEIGKVITWGRGFACVSPGENQLPVWLPTR  852
```

opioid binding protein/cell adehesion molecule, cow

```
cSRL-7d2-t      1    PPDITVNXGSSVTLLCLAIGRPEPT  75
                     ++|+||| ||||||||||||||||||
CPC4_BOVIN    142    SSDVTVNEGSSVTLLCLAIGRPEPT  166
```

env polyprotein - feline endogenous virus ECE1

```
cSRL-6g5-t      1    PLAKVVLQCXRALNMPYMEQGGYCMALREKCCFYTNHLGIIRDNMAMLK  147
                     +|+ |||| |+|++ ++++|| | ||+| ||||++| ||+||+|| |+
PIR:VCMVCE    533    SLLEVVLQNRRGLDLLFLQEGGLCAALKEECCFYADHTGIVRDSMAKLR  581
```

mitochondrial carnitine palmitoyltransferase II precursor, human

```
cSRL-2c2-t    226    SRRWFDKSFTFVVFKNGKMGLNAEHSWADAPIVAHLWEVSF  104
                     + |||||||  ++ |+|  +++ |||| |+ +| +  +  |
CPT2_HUMAN    348    TNRWFDKSFNLIIAKDGSTAVHFEHSWEDGVAVLRFFNEVF  388
```

trans-acting T-cell specific transcription factor GATA-3, mouse

```
cSRL-2g6-t    128    GGKALSGRTSSSAWKGACF*KDSIHTFCP  214
                     ||||||++ ++|+|+ +  | | |||   |
GAT3_MOUSE     98    GGKALSSHHTASPWNLSPFSKTSIHHGSP  126
```

FIG. 11B

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/06810

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :C12Q 1/68; C12P 19/34; C12N 15/00
US CL :435/6, 91.1, 91.3, 91.5; 935/26, 78, 80

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 91.1, 91.3, 91.5; 935/26, 78, 80

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, MEDLINE, BIOSIS
search terms: sequencing, genome, library, cosmid, multiplex

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Proc. Natl. Acad. Sci. USA, Volume 84, issued April 1987, Wahl et al., "Cosmid vectors for rapid genomic walking, restriction mapping, and gene transfer", pages 2160-2164, see entire document.	1-2 and 4-13
Y	Nature, Volume 325, issued 10 February 1987, Palca, "Interest in the human genome project reaches new heights", page 651, see entire document.	1-20
Y	Gene, Volume 26, issued 1983, Bates et al., "Double cos site vectors: simplified cosmid cloning", pages 137-146, see entire document.	11-13

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 04 AUGUST 1994	Date of mailing of the international search report AUG 18 1994
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer KENNETH R. HORLICK Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/06810

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Gene, Volume 79, issued 1989, Evans et al, "High efficiency vectors for cosmid microcloning and genomic analysis", pages 9-20, see entire document.	11-13
Y	Proc. Natl. Acad. Sci. USA, Volume 86, issued July 1989, Evans et al., "Physical mapping of complex genomes by cosmid multiplex anlysis", pages 5030-5034, see entire document.	3 and 14-20
Y,P	US, A, 5,219,726 (EVANS) 15 June 1993, see entire document.	3 and 14-20